

CYCLOPROPYLINDOLE DERIVATIVESTechnical Field

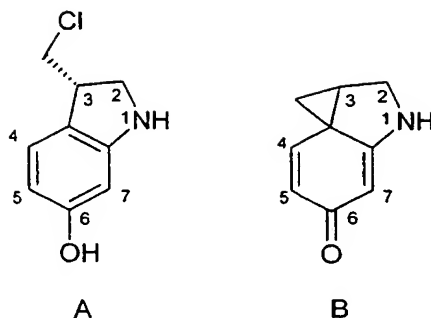
5 This invention relates to cyclopropylindoles (CPI) (which term is
used to encompass cyclopropylbenzindoles (CBI)) compounds and
their precursors, to methods of synthesizing these compounds on
solid supports, and to compounds of utility therein. This
invention further relates to collections of these compounds, and
10 methods for identifying and isolating CPI and precursor compounds
with useful and diverse activities from such collections.

Background to the invention

5 A large number of both synthetic and naturally occurring low
molecular weight ligands are known that interact with DNA via a
number of different mechanisms, including covalent or non-covalent
interaction in the minor or major grooves, intercalation between
base pairs or other types of non-specific interactions.

20 Of the class of ligands which interact with the minor groove, GC
specific ligands include Chromomycin, pyrrolo[2,1-
c][1,4]benzodiazepines (PBDs), Mitomycins and Ecteinasidins. Of
these, all but Chromomycin form a covalent bond with the DNA. Of
25 AT specific ligands, cyclopropylindoles form covalent bonds,
whilst compounds such as distamycin and netropsin do not.

30 Cyclopropylindole (CPI) compounds are a class of highly potent
antitumour antibiotics which includes the natural products CC-1065
(V.L. Reynolds et al, J. Antibiot., 39, 1986, 319-314) and the
duocarmycins (D.L. Boger, Pure & Appl. Chem., 66, 1994, 837-844),
having IC₅₀s in the low pM range in tumour cells growing *in vitro*.
They are of the general structures A and B:



Studies with compounds that model the binding subunit have shown that the more stable open chain seco-precursors (e.g. A) are as potent as the cyclopropylindole compounds (e.g. B). Further, ring closure is not essential for DNA binding.

A number of synthetic analogues of the natural products have been prepared in which the oxygen of A is protected as a carbamate that must be cleaved (by non-specific enzymatic hydrolysis) for activity. Further analogues of a similar type are disclosed in WO88/04659 and WO91/16324. Analogues where the 6 substituent is N or S are disclosed in WO 97/07097 and WO 98/11101.

Compounds having biological activity can be identified by screening collections of compounds (i.e. libraries of compounds) produced through synthetic chemical techniques. Such screening methods include methods wherein the library comprises a plurality of compounds synthesized at specific locations on the surface of a solid support where a receptor is appropriately labelled to identify binding to the compound, e.g., fluorescent or radioactive labels. Correlation of the labelled receptor bound to the support with its location on the support identifies the binding compound (US 5,143,854).

Central to these methods is the screening of a multiplicity of compounds in the library and the ability to identify the structures of the compounds which have a requisite biological activity. In order to facilitate synthesis and identification, the compounds in the library are typically formed on solid supports. Usually each such compound is covalently attached to

the support via a cleavable or non-cleavable linking arm. The libraries of compounds can be screened either on the solid support or as cleaved products to identify compounds having good biological activity.

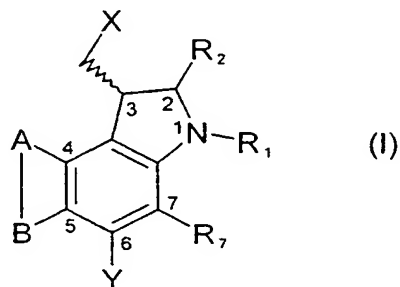
5

Disclosure of the Invention

The present invention provides CPI compounds with structures that allow them to be joined to combinatorial chains, as well as combinatorial libraries containing CPIs themselves.

10

A first aspect of the present invention relates to compounds of formula I:



15

wherein X is an electrophilic leaving group;

Y is selected from NH-Prot, O-Prot, S-Prot, NO₂, NHOH, N₃, NHR, NRR, N=NR, N(O)RR, NHSO₂R, N=NPhR, SR or SSR, where Prot

20

represents a protecting group;

A and B collectively represent a fused benzene or pyrrole ring (in either orientation), which is optionally substituted by up to respectively 4 or 2 groups independently selected from R, OH, OR, halo, nitro, amino, Me₃Sn, CO₂H, CO₂R;

25

R₁ is a nitrogen protecting group, where if Y includes a protecting group, these protecting groups are orthogonal;

R₂ and R₇ are independently selected from H, R, OH, OR, halo, nitro, amino, Me₃Sn;

wherein R is selected from

30

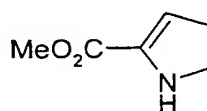
- (a) a lower alkyl group having 1 to 10 carbon atoms,
- (b) an aralkyl group (i.e. an alkyl group with one or more aryl substituents), preferably of up to 12 carbon atoms;

100699202 022202

the alkyl group of (a) or (b) optionally containing one or more carbon-carbon double or triple bonds, which may form part of a conjugated system; and

(c) an aryl group, preferably of up to 12 carbon atoms; and

5 wherein R is optionally substituted by one or more halo, hydroxy, amino, or nitro groups, and optionally contains one or more hetero atoms, which may form part of, or be, a functional group; except that when R₁ is Boc, Y is NO₂, X is Cl, and R₂ and R₇ are H, then A and B do not collectively represent either an unsubstituted
10 benzene ring or:

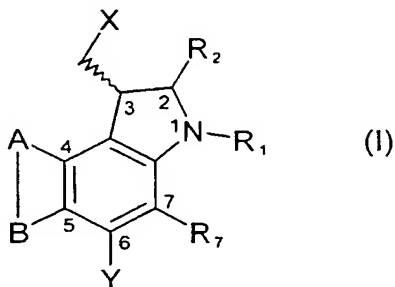


100699202 "022202
15 If R is an aryl group, and contains a hetero atom, then R is a heterocyclic group. If R is an alkyl chain, and contains a hetero atom, the hetero atom may be located anywhere in the alkyl chain, e.g. -O-C₂H₅, -CH₂-S-CH₃, or may form part of, or be, a functional group, e.g. carbonyl, hydroxy.

20 R is preferably independently selected from a lower alkyl group having 1 to 10 carbon atoms, or an aralkyl group, preferably of up to 12 carbon atoms, or an aryl group, preferably of up to 12 carbon atoms, optionally substituted by one or more halo, hydroxy, amino, or nitro groups. It is more preferred that R is
25 independently selected from lower alkyl groups having 1 to 10 carbon atoms optionally substituted by one or more halo, hydroxy, amino, or nitro groups. It is particularly preferred that R is an unsubstituted straight or branched chain alkyl group, having 1 to 10, preferably 1 to 6, and more preferably 1 to 4, carbon atoms,
30 e.g. methyl, ethyl, propyl, butyl.

Y is preferably NH-Prot, O-Prot, S-Prot, and more preferably NH-Prot.

These compounds are useful in the synthesis of collections of CBI and CPI precursors. Compounds of formula I:



wherein X, A, B, R₁, R₂, R₇ are as defined above, and Y is selected from NH₂, NH-Prot, OH, O-Prot, SH, S-Prot, NO₂, NHOH, N₃, NHR, NRR, N=NR, N(O)RR, NHSO₂R, N=NPhR, SR or SSR, where Prot represents a protecting group can be attached to a solid support, e.g. via a connecting link which may comprise a chain of combinatorial units. This is a second aspect of the invention, i.e. the use of these compounds in methods of combinatorial chemistry synthesis, wherein the compound is joined to a solid support by a chain comprising at least one combinatorial unit. The preferences for R expressed in the first aspect apply to this aspect of the invention.

Y in this second aspect is preferably NH₂, NH-Prot, OH, O-Prot, SH, S-Prot, more preferably NH₂, NH-Prot, SH, S-Prot, and most preferably NH-Prot. As an alternative, OH and O-Prot are preferred.

Furthermore, compounds of formula (I) with substituents as defined in this aspect, where the fused ring, represented by -A-B-, bears a substituent -CO₂H or -CO₂R, may be used as combinatorial units (see below).

The term 'protecting group' (and more specifically 'nitrogen protecting group') has the meaning usual in synthetic chemistry, particularly for 'nitrogen protecting group', in synthetic peptide chemistry. It means any group which may be covalently bound to the protected atom of the CBI or CPI grouping, and permits reactions to be carried out upon the molecule containing this grouping without its removal. Nevertheless, it is able to be removed from the protected atom without affecting the remainder of

the molecule. Suitable nitrogen protecting groups for the present invention include Fmoc (9-fluorenylmethoxycarbonyl), Nvoc (6-nitroveratryloxycarbonyl), Teoc (2-trimethylsilylethyloxycarbonyl), Troc (2,2,2-trichloroethyloxycarbonyl), Boc (t-butyloxycarbonyl), CBZ (benzyloxycarbonyl), Alloc (allyloxycarbonyl) or Psec (2(-phenylsulphonyl)ethyloxycarbonyl). Suitable oxygen (hydroxyl) protecting groups include t-butyl ethers, Benzyl ethers, Silyl ethers, MOM (methoxy methyl ethers), MEM (2-methoxy ethoxy methyl ethers) or acetates. Suitable sulphur (thiol) protecting groups include benzyl, nitrogenyl thioether or fluorenylmethyl thioethers. Other suitable groups are described in Protective Groups in Organic Synthesis, T Green and P Wuts, published by Wiley, 1991, which is incorporated herein by reference.

In the first and second aspects of the invention it is preferred that the nitrogen protecting group has a carbamate functionality where it binds to the nitrogen atom to be protected.

The term 'orthogonal' in relation to 'protecting groups' has the meaning usual in synthetic chemistry. It means that one protecting group may be selectively removed without affecting the other protecting group. This is achieved by using protecting groups which are sensitive to different removal conditions.

Examples of orthogonal protecting group pairs are:

CBZ carbamate and Boc Carbamate - the CBZ group can be removed by hydrogenation whilst the BOC carbamate remains intact.

Alternatively, the BOC group can be removed with TFA whilst the CBZ group remains intact. The same applies to a CBZ carbamate and a tert-butyl ether.

Fmoc carbamate and t-butyl ether - The Fmoc group can be removed with base (50% piperidine in DMF) without affecting the t-butyl protection. The t-butyl side chain protecting group is removed by acid (TFA, trifluoroacetic acid). If necessary t-butyl ethers can be cleaved in the presence of Fmoc as Fmoc is unaffected by acid.

Alloc carbamate and Silyl ethers - an allyl carbamate is cleaved specifically by zero valent palladium catalysts, Silyl ethers are cleaved with fluoride ions or dilute acid (neither reagent affects Alloc and palladium does not affect Silyl ethers).

5

An example of non-orthogonal pair includes CBZ and Fmoc, as both are cleaved by hydrogenation.

10

An electrophilic leaving group is a group that is readily eliminated from the molecule and carries with it an electron-pair. These may be termed "nucleofugal" leaving groups.

15

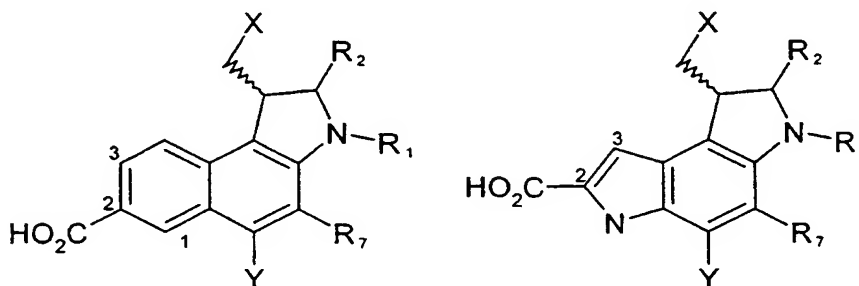
In the present invention, it is preferred that X is either halogen or OSO_2R , where R is as defined earlier. A halogen group means a fluoro, chloro, bromo or iodo group. It is more preferred that X is Cl.

20

If X is OSO_2R , then R is preferably $-\text{CH}_3$ (mesylate), $-(p\text{-Me})\text{Ph}$ (tosylate), $-\text{CF}_3$ (triflate) or $-\text{C}_4\text{F}_9$ (nonaflate)

25

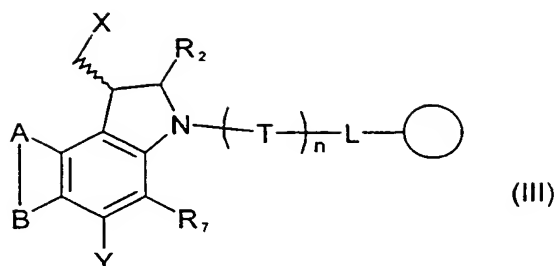
In the present invention it is also further preferred that the 4,5 fused ring is substituted by $-\text{CO}_2\text{R}$ in the 2 or 3 position if it is a benzene ring, and the 2 position if it is a pyrrole ring, e.g.



where R_2 and R_7 are preferably H, with the proviso that Y is not NO_2 when R_1 is Boc, R_2 and R_7 are H, and X is Cl.

30

A third aspect of the present invention relates to compounds of the formula **III**:

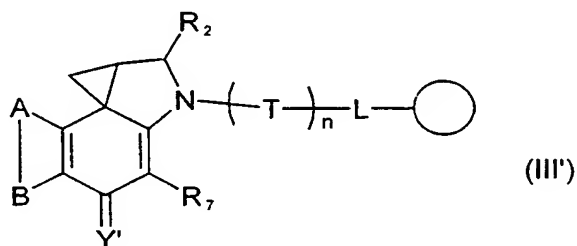


wherein X, Y, A, B, R₂ and R₇ are as defined in the second aspect;
T is a combinatorial unit;

5 and n is a positive integer, where if n is greater than 1, each T
may be different;

L is a linking group, or less preferably a single bond;
and O is a solid support.

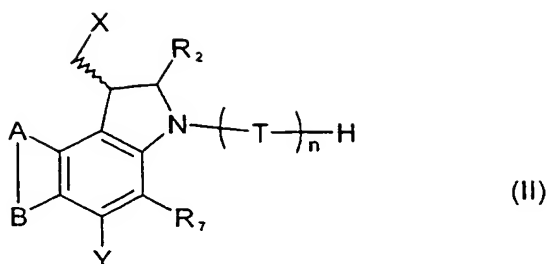
10 A fourth aspect of the present invention relates to compounds of
formula III':



15 where A, B, R₂, R₇, T, n, L and O are as defined in the third
aspect; and

Y' is NH, NR, O or S, preferably NH or S.

Compounds of this aspect are obtainable from compounds of the
second aspect by removal of the protecting group preferably
20 carbamate in Y (if present), or by other appropriate reactions,
e.g. reduction and/or basic conditions. The cyclisation may also
occur spontaneously. A fifth aspect of the present invention
relates to compounds of formula II:



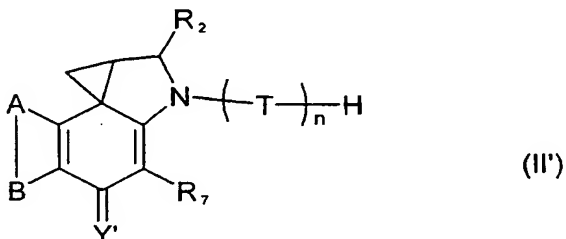
wherein X, Y, A, B, R₂, R₇, T and n are as defined in the third aspect of the invention.

5

Compounds of this aspect are obtainable by cleavage of the linking group of the appropriate compound of the third aspect.

A sixth aspect of the present invention relates to compounds of the formula II':

10



wherein A, B, T, n, R₂ and R₇ are as defined in the fifth aspect of the invention; and

15

Y' is NH, NR, O, or S, preferably NH or S.

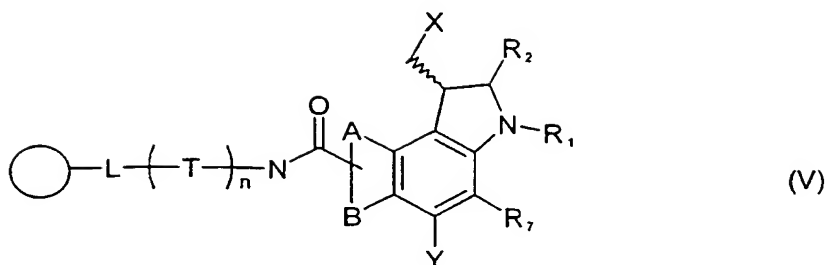
Compounds of this aspect are obtainable by cleavage of the linking group of the appropriate compound of the third aspect (see below).

20

Alternatively, they are obtainable from compounds of the fourth aspect by removal of the protecting group in Y (if present), or by other appropriate reactions, e.g. reduction and/or basic conditions. The cyclisation may also occur spontaneously.

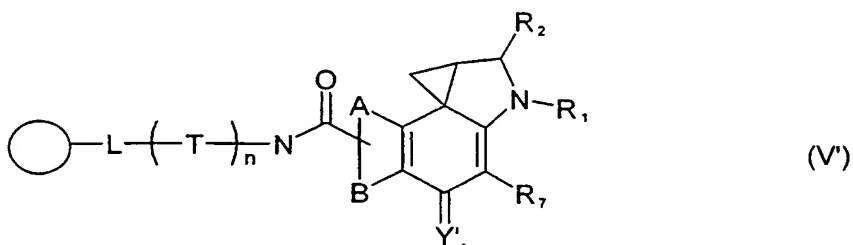
25

A seventh aspect of the present invention relates to compounds of formula V:



where A, B, Y, R₁, R₂, R₇, and X are defined in the second aspect of the invention, and T, n, L and O are as defined in the third aspect of the present invention.

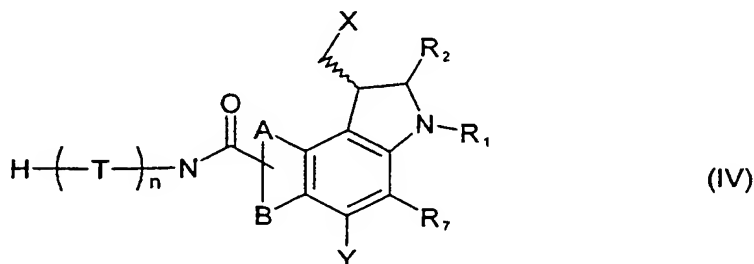
An eighth aspect of the present invention relates to compounds of formula V':



where A, B, R₁, R₂, R₇, T, n, L and O are as defined in the seventh aspect; and Y' is NH, NR, O or S, preferably NH or S.

Compounds of this aspect are obtainable from compounds of the seventh aspect by removal of the protecting group in Y (if present), or by other appropriate reactions, e.g. reduction and/or basic conditions. The cyclisation may also occur spontaneously.

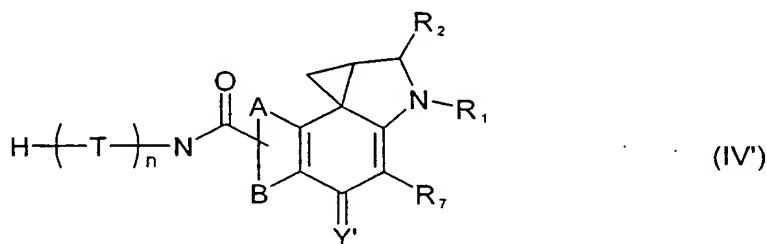
A ninth aspect of the present invention relates to compounds of formula IV:



wherein A, B, X, Y, T, n, R₁, R₂ and R₇ are as defined in the seventh aspect of the invention.

Compounds of this aspect are obtainable by cleavage of the linking group of the appropriate compound of the seventh aspect.

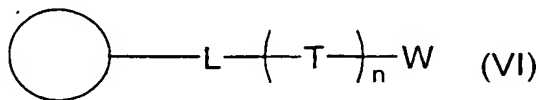
A tenth aspect of the present invention relates to compounds of formula IV' :



wherein A, B, T, n, R₁, R₂ and R₇ are as defined in the ninth aspect of the invention; and Y' is NH, NR, O, or S, preferably NH or S.

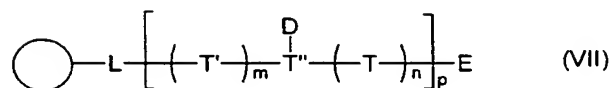
15 Compounds of this aspect are obtainable by cleavage of the linking group of the appropriate compound of the eighth aspect (see below). Alternatively, they are obtainable from compounds of the ninth aspect by removal of the protecting group in Y (if present), or by other appropriate reactions, e.g. by reduction and/or basic conditions. The cyclisation may also occur spontaneously.

20 An eleventh aspect of the invention is the preparation of compounds according to either the third or seventh aspects by reaction of a compound of formula VI:



30 with a compound of formula I according to the second aspect, where T, n, L and O are as defined in these aspects, and W is H or an atom or group for providing a functional group capable of reaction with -COOH or -NH₂. This reaction will include the necessary protection and deprotection steps so as to selectively achieve a compound according to the third or seventh aspect.

A twelfth aspect of the invention relates to compounds of formula VII:



5 wherein:

O, T, and L are as defined in the third aspect of the invention;

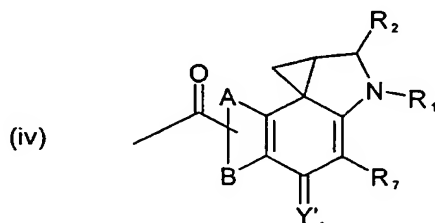
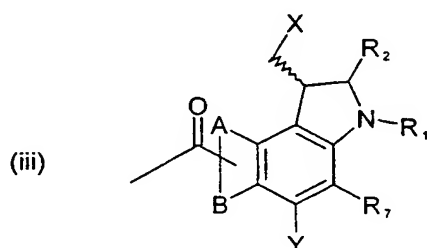
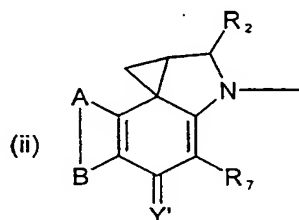
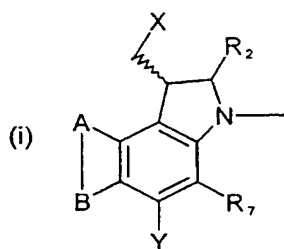
n and m are positive integers, or one of them may be zero;

T' is a combinatorial unit, where each T' may be different if m is greater than 1;

10 T'' is a combinatorial unit which provides a site for the attachment of D;

D is selected from:

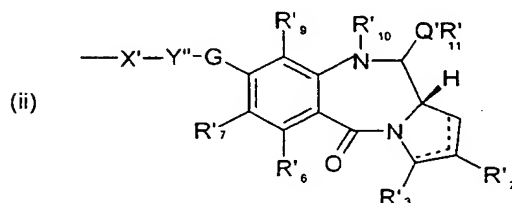
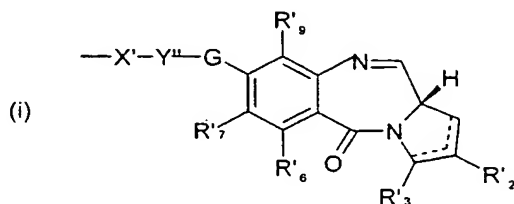
(a)



10069202.02202

wherein A, B, Y, Y', R₂ and R₇ are as defined in the second or third aspects of the invention; or

(b)



- 5 wherein X' is selected from CO, NH, S, or O;
 G is O, S, NH, or a single bond;
 R'₂ and R'₃ are independently selected from: H, R, OH, OR, =O, =CH-R, =CH₂, CH₂-CO₂R, CH₂-CO₂H, CH₂-SO₂R, O-SO₂R, CO₂R, COR and CN, and there is optionally a double bond between C₂ and C₃;
 10 R'₆, R'₇, and R'₉ are independently selected from H, R, OH, OR, halo, nitro, amino, Me₃Sn;
 R'₁₁ is either H or R;
 Q' is S, O or NH;
 R'₁₀ is a nitrogen protecting group;
 15 Y'' is a divalent group such that HY = R;
 p is a positive integer, where if p is greater than 1, for each repeating unit, the meaning of T, T', T'' and D and the values of n and m are independently selected; and
 E is selected from the same possibilities as D, provided that at
 20 least one group D or E is selected from (a).

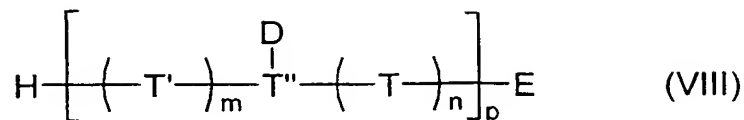
If, for example, D terminates with CO then the site on T'' may be NH, and if D terminates with NH, S or O, then the site on T'' may be CO.

25

In preferred embodiments of this aspect, one of D and E is selected from (a) and the other is selected from (b).

10069202-022202

A thirteenth aspect of the invention relates to compounds of formula (VIII):

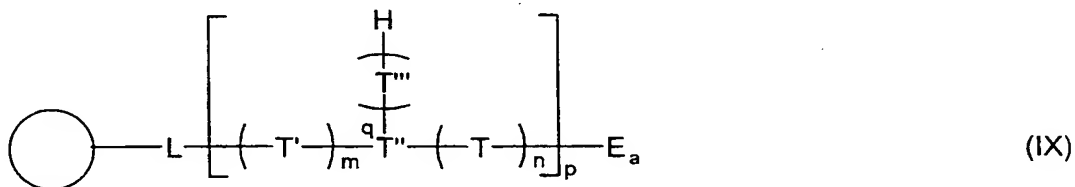


- 5 wherein L, T, T', T'', D, E, n, m and p are as defined in the twelfth aspect of the invention.

In preferred embodiments of this aspect, one of D and E is selected from (a) and the other is selected from (b).

10

A fourteenth aspect of the invention relates to compounds of formula (IX):



15

wherein O, L, T, T', T'', n, m and p are as defined in the twelfth aspect of the invention;

T''' is a combinatorial unit;

q is a positive integer, where if q is greater than 1, each T'''

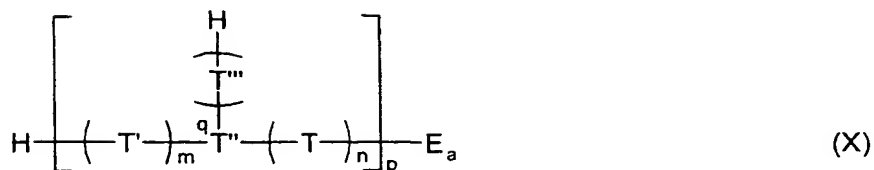
- 20 may be different; and

E_a is selected from the group (a) of E as defined in the twelfth aspect of the invention;

where if p is greater than 1, for each repeating unit the meaning of T, T', T'', T''' and the values of n, m and q are independently

- 25 selected.

A fifteenth aspect of the invention relates to compounds of formula (X):



wherein L, T, T', T'', T''', E_a, n, m, p and q are as defined in the
 5 fourteenth aspect of the invention.

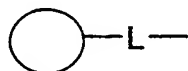
It is recognised that the compounds of formulae I, II, II', III, III', IV, IV', V, V', VII, VIII, IX and X may exist in different enantiomeric or diastereomeric forms. In such cases it is to be
 10 understood that the above formulae represent any possible enantiomeric or diastereomeric form or a mixture thereof.

Solid support

15 The term 'solid support' refers to a material having a rigid or semi-rigid surface which contains or can be derivatized to contain reactive functionalities which can serve for covalently linking a compound to the surface thereof. Such materials are well known in the art and include, by way of example, silicon dioxide supports
 20 containing reactive Si-OH groups, polyacrylamide supports, polystyrene supports, polyethyleneglycol supports, and the like. Such supports will preferably take the form of small beads, pins/crowns, laminar surfaces, pellets, disks. Other conventional forms may also be used.

Linker group

25 The linking groups suitable for the present application are ones which usually provide in the structure:
 30

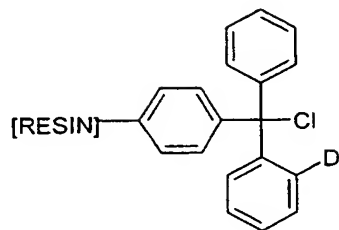


at least one covalent bond which can be readily broken by specific chemical reactions, such as oxidation (e.g. using DDQ or CAN), nucleophilic attack (e.g. with an amine on an oxime linker, or an

alkylated sulfamyl linker (or by light, e.g. 365 nm, or changes in pH, e.g. by adding TFA) thereby providing for liberation of compounds free from the solid support. The methods employed to break the covalent bond are selected so as to be specific for the desired bond breakage thereby preventing unintended reactions from occurring elsewhere on the complex. The linking group is selected relative to the synthesis of the compounds to be formed on the solid support so as to prevent premature cleavage of this compound from the solid support as well as to limit interference by any of the procedures employed during compound synthesis on the support.

Examples of resins incorporating cleavable linking groups are set out in the table below, which also indicates the groups that can be immobilised thereon, along with the suggested cleavage methods for the linking group. Such resins are commercially available (e.g. from NovaBiochem).

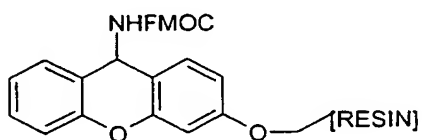
Linker/Resin Type	Immobilises	Cleavage Method
2-Chlorotrityl chloride	RNH ₂ , RCO ₂ H, ROH, RSH	1-50% TFA
Trityl chloride	RNH ₂ , RCO ₂ H, ROH, RSH	1-5% TFA
2-Methoxytrityl chloride	RNH ₂ , RCO ₂ H, ROH, RSH	1-5%
Rink amide resin	RCO ₂ H	95% TFA
Sieber amide resin	RCO ₂ H	1% TFA
4-Sulfamyl- benzoyl	RCO ₂ H	Alkylation /amines
Wang resin	ROH, ArOH, RNH ₂ , RCO ₂ H	15-95% TFA or DDQ or CAN
HMPB-BHA	ROH, ArOH, RCO ₂ H	1% TFA
Bromoethyl photolinker	RNH ₂ , RCO ₂ H, ROH, RSH	hν
Hydroxy ethyl photolinker	RCO ₂ H	hν
Aminoethyl photolinker	RCO ₂ H	hν

Structures

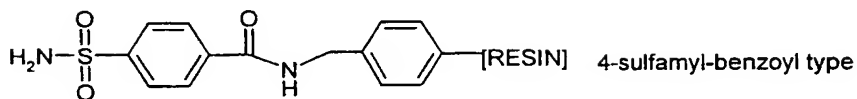
D = Cl: 2-chlorotrityl chloride type
 D = H: trityl chloride type
 D = OMe: 2-methoxytrityl chloride



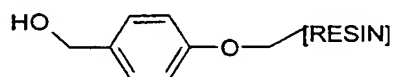
Rink amide type



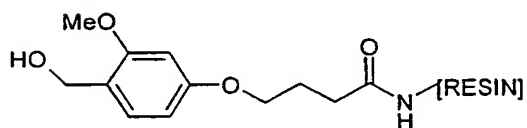
Sieber amide type



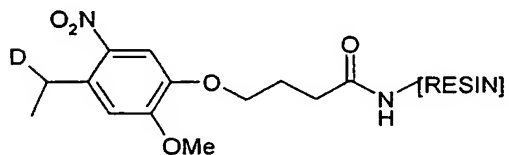
4-sulfamyl-benzoyl type



Wang type



HMPB-BHA type



D = NH₂: amino-ethyl type
 D = OH: hydroxy-ethyl type
 D = Br: bromo ethyl type

20220220" 20226900T

For CPI precursors the most preferred linking group is one which may be deemed photolytically cleavable. Further, the Rink amide linker is particularly suitable.

5

It is also possible that the linking group is a simple functionality provided on the solid support, e.g. amine, and in this case the linking group may be not be readily cleavable. This type of linking group is useful in the synthesis of large split and mix libraries which will be subjected to on-bead screening (see below), where cleavage is unnecessary. Such resins are commercially available from a large number of companies including NovaBiochem, Advanced ChemTech and Rapp Polymere. These resins include amino-Tentagel and aminomethylated polystyrene resin.

10

15

Combinatorial Unit

20

25

30

35

The term 'combinatorial unit' means any monomer unit which can be used to build a chain attached to the solid support, usually by a linking group. Usually combinatorial units will comprise at least two different functional groups to provide them with this chain building ability. For example, amino acids comprise both carboxylic acid and amine moieties. Sometimes the combinatorial unit may require more than two functionalities, e.g. if it has to bond to a further moiety. Examples of molecules suitable for such chain building are found in Schreiber et al. (JACS, 120. 1998, pp.23-29), which is incorporated herein by reference. An important example of a unit is an amino acid residue. Chains may be synthesised by means of amine-protected amino acids. Fmoc protected amino-acids are available from a number of sources, such as Sigma and Nova Biochem. Both natural and unnatural amino acids can be used, e.g. D- and L-amino acids and heterocyclic amino acids. In particular, heterocyclic amino acids of the type found in the construction of netropsin and distamycin are of interest because of their DNA-recognition properties.

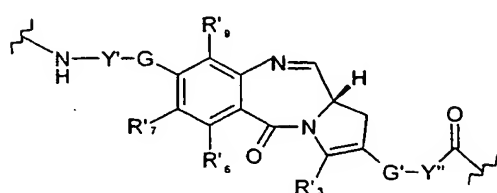
Amine units can be used to make up peptoids: see Soth, M.J. and Nowick, J.S. 1997, Unnatural oligomer libraries, *Curr. Opin. Chem.*

Biol. 1, no. 1: 120-129; Zuckermann et al., 1994, Discovery of Nanomolecular Ligands for 7-Transmembrane G-Protein-Coupled Receptors from a Diverse N-(Substituted)glycine Peptoid Library, *Journal of Medicinal Chemistry* 37: 2678-85; Figliozzi, GMR et al., 1996, Synthesis of N-substituted Glycine Peptoid Libraries, *Methods in Enzymology*, 267: 437-47; Simon, R J et al., 1992, Peptoids: A Modular Approach to Drug Discovery, *Proc. Natl. Acad. Sci. USA*, 89:9367-71; which are all incorporated herein by reference.

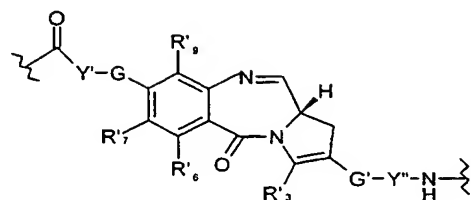
Other combinatorial units include PNAs (peptidonucleic acids): PE Nielsen, et al, *Science*, 1991, 254, 1497; M Egholm, et al, *Nature*, 1993, 365, 566; M Egholm et al, *JACS*, 1992, 114, 1895; S C Brown, et al, *Science*, 1994, 265, 777; 5. K Saha, et al, *JOC*, 1993, 58, 7827; oligoureas: Burgess K, et al, 1995, Solid Phase Synthesis of Unnatural Biopolymers Containing Repeating Urea Units. *Agnew. Chem. Int. Ed. Engl* 34, no. 8:907; Burgess K, et al, 1997, Solid Phase Synthesis of Oligoureas; *Journal of the American Chemical Society* 119: 1556-64; and oligocarbamates: Moran E J et al, 1995, Novel Biopolymers for Drug Discovery. *Biopolymers (Peptide Science)*; *John Wiley and Sons* 37: 213-19; Cho C Y et al, 1993, An Unnatural Biopolymer. *Science* 261: 1303-5; Paikoff S F et al, 1996, The Solid Phase Synthesis of N-Alkylcarbamate Oligomers. *Tetrahedron Letters* 37, no. 32: 5653-56. All of these documents are incorporated herein by reference.

A further aspect of the present invention relates to combinatorial units having formula I, where the fused ring collectively represented by A and B is substituted by CO₂R.

Further combinatorial units of relevance to this invention are those of formulae (XIa/b):



(XIa)



(XIb)

wherein R'_3 , R'_6 , R'_7 , R'_9 , G and Y' are as defined in the twelfth aspect of the invention, and G' and Y'' are independently selected from the possible groups for G and Y' respectively. In order to synthesise combinatorial chains containing such combinatorial units, the units may need to be joined to the chain in their protected form (see definition of groups D(b) above). It is possible that the combinatorial units may remain in their protected form until the compound is cleaved from the solid support, or until the other components in the compound are deprotected.

Combinatorial units may also be of formula I with substituents as defined in the second aspect of the invention, where the fused ring, represented by $-A-B-$, bears a substituent which is $-CO_2H$ or $-CO_2R$.

The present invention relates to libraries, or collections, of compounds, all of which are represented by a single one of the formulae II, II', III, III', IV, IV', V, V', VII, VIII, IX or X. The diversity of the compounds in a library may reflect the presence of compounds differing in the identities of one or more of (as appropriate) X , Y , Y' , A , B , R_2 , R_7 , X' , Y'' , G , R'_2 , R'_3 , R'_6 , R'_7 , R'_9 , R'_{10} , Q' , R'_{11} and/or in the identities of the combinatorial units T , T' , T'' and T''' (when present). The number of members in the library depends on the number of variants, and the number of possibilities for each variant. For example, if it is the combinatorial units which are varied, and there are 3 combinatorial units, with 3 possibilities for each unit, the library will have 27 compounds. 4 combinatorial units and 5 possibilities for each unit gives a library of 625 compounds. If,

for instance, there is a chain of 5 combinatorial units with 17 possibilities for each unit, the total number of members in the library would be 1.4 million. A library may therefore comprise more than 1 000, 5 000, 10 000, 100 000 or a million compounds, which may be arranged as described below.

In the case of free compounds (formulae II, II', IV, IV', VIII, X) the individual compounds are preferably in discrete volumes of solvents, e.g. in tubes or wells. In the case of bound compounds (formulae III, III', V, V', VII, IX) the individual compounds are preferably bound at discrete locations, e.g. on respective pins/crowns or beads. The library of compounds may be provided on a plate which is of a suitable size for the library, or may be on a number of plates of a standard size, e.g. 96 well plates. If the number of members of the library is large, it is preferable that each well on a plate contains a number of related compounds from the library, e.g. from 10 to 100. One possibility for this type of grouping of compounds is where only a subset of the combinatorial units, or substituents, are known and the remainder are randomised; this arrangement is useful in iterative screening processes (see below). The library may be presented in other forms that are well-known.

A further aspect of the present invention is a method of preparing a diverse collection, or library of compounds, as discussed above.

If the diversity of the library is in the combinatorial units, then the library may be synthesised by the stepwise addition of protected combinatorial units to a CPI/CBI precursor core, each step being interposed by a deprotection step. Such a method is exemplified later. Libraries of this type can be prepared by the method known as "split and mix" which is described in Furka, A; Sebestyen, F; Asgedom, M and Dibo, G; General Method of Rapid Synthesis of Multicomponent Peptide Mixtures; International Journal of Peptide and Protein Research; 1991, 37, 487-193, which is incorporated herein by reference. If the diversity of the library is in the substituent groups, the library may be synthesised by carrying out the same synthetic methods on a variety of starting materials or key intermediates, which already possess the necessary substituent patterns.

The present invention also relates to a method of screening the compounds of formula II, II', III, III', IV, IV', V, V', VII, VIII, IX or X to discover biologically active compounds. The screening can be used to assess the binding interaction with nucleic acids, e.g. DNA or RNA, or proteins, or to assess the affect of the compounds against protein-protein or nucleic acid-protein interactions, e.g. transcription factor DP-1 with E2F-1, or estrogen response element (ERE) with human estrogen receptor (a 66 kd protein which functions as a hormone-activated transcription factor, the sequence of which is published in the art and is generally available). The screening may also be used to assess the cytotoxicity of the compounds against a variety of cell lines. The screening can be carried out by bringing the target macromolecules into contact with individual compounds or the arrays or libraries of individual compounds described above, and selecting those compounds, or wells with mixtures of compounds, which show the strongest effect.

This effect may simply be the cytotoxicity of the compounds in question against cells or the binding of the compounds to nucleic acids. In the case of protein-protein or nucleic acid-protein interactions, the effect may be the disruption of the interaction studied.

The binding of the compounds to nucleic acids may be assessed by labelling oligomers which contain a target sequence, and measuring the amount of labelled oligomers that bind to the compounds tested. The labelling may either be radiolabelling, or alternatively be labels detectable under visible or ultra-violet light. If this latter form of screening is carried out on compounds bound to solid supports which are in separate locations, the screening for results can be carried out visually under a microscope. A similar technique is described in detail in DNA-Binding ligands from peptide libraries containing unnatural amino acids, Lescrinier et al., *Chem Eur J*, 1998, 425-433. These techniques are particularly suited to a one-step screening of a

complete library of compounds, especially a large library made by the "split and mix" method described above.

Protein-protein interactions can be measured in a number of ways, e.g. FRET (fluorescence resonance energy transfer) which involves labelling one of the proteins with a fluorescent donor moiety and the other with an acceptor which is capable of absorbing the emission from the donor; the fluorescence signal of the donor will be altered depending on the interaction between the two proteins. Another method of measuring protein-protein interactions is by enzymatic labelling, using, for example, horseradish peroxidase.

The screening process may undergo several iterations by selecting the most active compounds, or group of compounds, tested in each iteration; this is particularly useful when testing arrays of wells which include mixtures of related compounds. Furthermore, if the wells contain compounds for which only a subset of the combinatorial units, or substituents, are known, but the rest are randomised, subsequent iterations can be carried out by synthesising compounds possessing the selected known (and successful) combinatorial unit, or substituent, pattern, but with further specified combinatorial units, or substituents, replacing the previously randomised combinatorial units, or substituents, adjacent the already known pattern; the remaining combinatorial units, or substituents, are randomised as in the previous iteration. This iterative method enables the identification of active members of large libraries without the need to isolate every member of the library.

A further feature of this aspect is formulation of selected compound or compounds with pharmaceutically acceptable carriers or diluents.

In yet further aspects, the invention provides a pharmaceutical composition comprising a compound of formula II, II', IV, IV', VIII or X and a pharmaceutically acceptable carrier or diluent; and the use of a compound of formula II, II', IV, IV', VIII or X in the manufacture of a medicament for the treatment of a

gene-based disease, or a bacterial, parasitic or viral infection. Gene-based disease include neoplastic disease and, for example, Alzheimer's disease.

- 5 Compounds of formula II, II', IV, IV', VIII or X may be used in a method of therapy against a gene-based disease, such as cancer or Alzheimer's disease, or a viral, parasitic or bacterial infection.

- 10 Another aspect of the present invention relates to the use of compounds of formula III, III', V, V', VII or IX in diagnostic methods. A compound of formula III, III', V, V', VII or IX which binds to an identified sequence of DNA or a protein known to be an indicator of a medical condition can be used in a method of diagnosis. The method may involve passing a sample, e.g. of
- 15 appropriately treated blood or tissue extract, over an immobilised compound of formula III, III', V, V', VII or IX, for example in a column, and subsequently determining whether any binding of target DNA to the compound of formula III, III', V, V', VII or IX has taken place. Such a determination could be carried out by passing
- 20 a known amount of labelled target DNA known to bind to compound III, III', V, V', VII or IX through the column, and calculating the amount of compound III, III', V, V', VII or IX that has remained unbound.

- 25 A further aspect of the present invention relates to the use of compounds of formula II, II', IV, IV', VIII or X in target validation. Target validation is the disruption of an identified DNA sequence to ascertain the function of the sequence, and a compound of formula II, II', IV, IV', VIII or X can be used to
- 30 selectively bind an identified sequence, and thus disrupt its function.

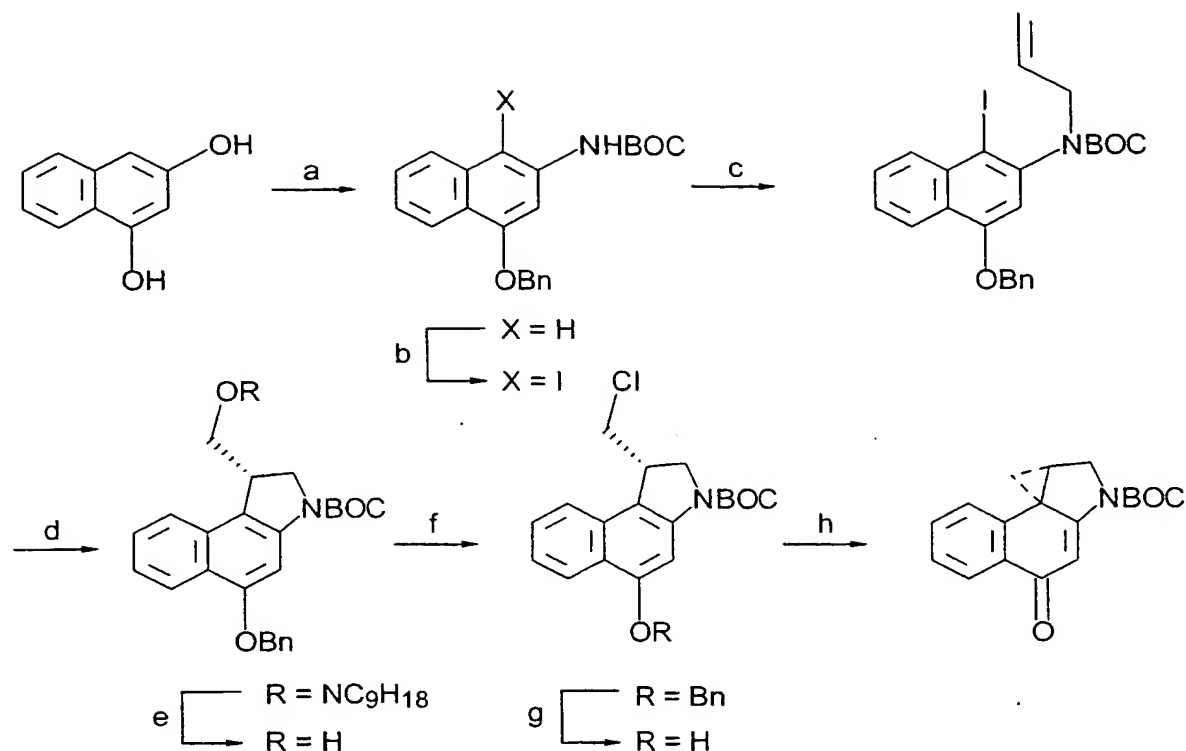
- Another aspect of the present invention relates to the use of compounds of formulae II, II', IV, IV', VIII or X in functional
- 35 genomics to ascertain the biological function of genes, by blocking this biological action.

Preferred Synthetic Strategies

Compounds of formula I can be synthesised by applying the methods described below. A review of methods of synthesising CPIs was carried out by Boger (Boger, D.L. et al., *J.A. Chem. Rev.* **1997**, 97, 787-828)

Boger Synthesis of N-BOC-CBI

Selective C4 iodination of N-BOC-4-(benzyloxy)naphthylamine, readily accessible in three steps from the commercially available 1,3-dihydroxynaphthalene, followed by N-alkylation with allyl bromide provided the required substrate for the 5-exo-trig aryl radical-alkene cyclization (Boger, D.L. et al., *O. J. Am. Chem. Soc.* **1989**, 111, 6461-6463. Boger, D.L. et al., *O. J. Org. Chem.* **1990**, 55, 5823-5832) (Scheme 1). Treatment with Bu₃SnH-TEMPO (Boger, D.L.; McKie, J.A. *J. Org. Chem.* **1995**, 3, 1429-1453) and subsequent reduction with Zn afforded the 3-(hydroxymethyl)indoline CBI precursor. Conversion to the primary chloride and catalytic hydrogenolysis of the benzyl ether, followed by direct resolution on a semipreparative chiral HPLC column afforded both enantiomers (Boger, D.L.; Yun, W. *J. Am. Chem. Soc.* **1994**, 116, 7996-8006). Subsequent spirocyclization completed the synthesis.

Scheme 1

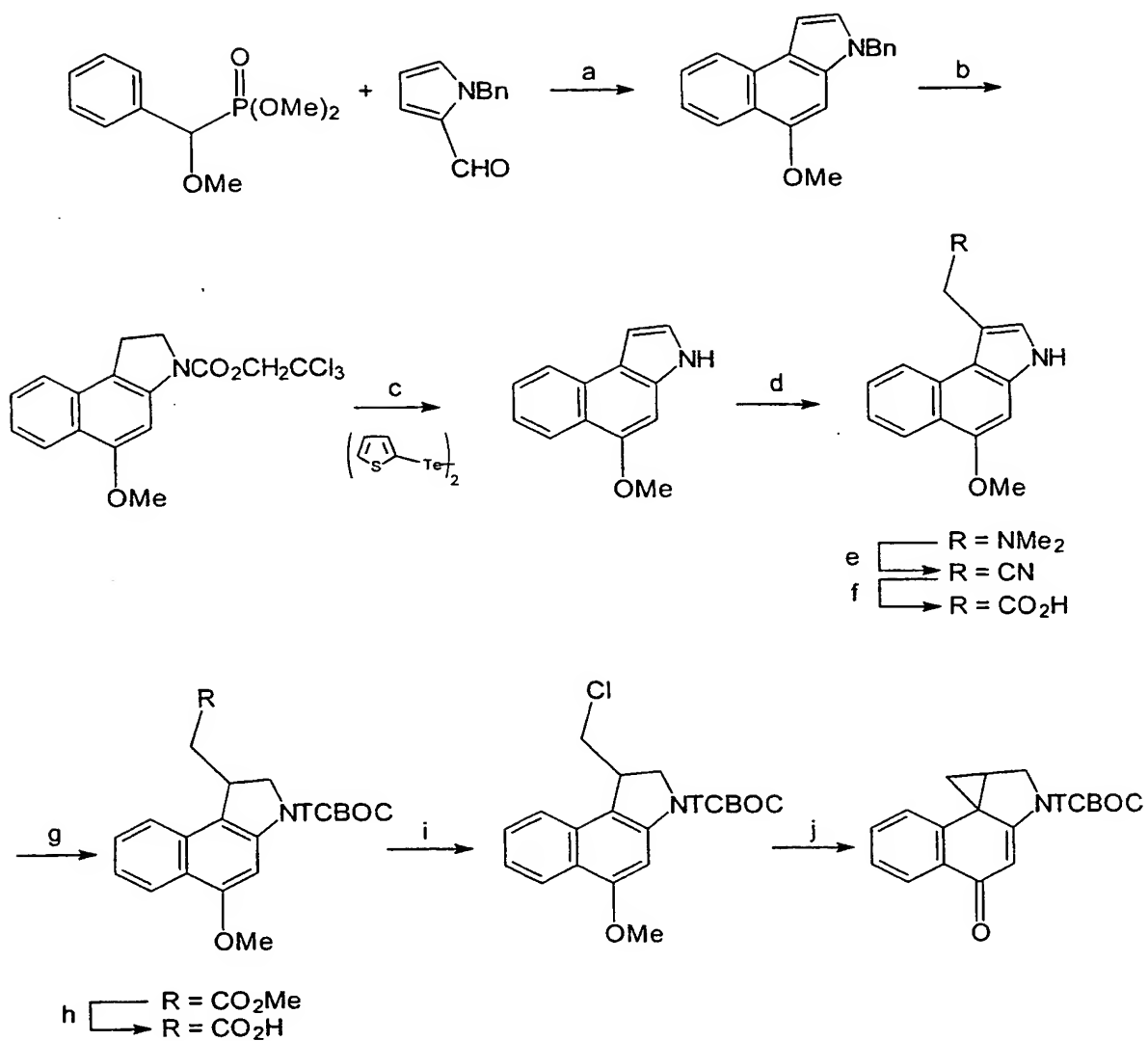
5

Reagents and conditions: a: NH_3 , BOC_2O ; BnBr ; b: NIS ; c: AllylBr , NaH ; d: Bu_3SnH , TEMPO ; e: Zn ; f: $\text{Ph}_3\text{P-CCl}_4$; g: H_2 , Pd-C ; h: NaH

Cava synthesis of CBI

- 10 Cava's route to CBI uses a heterocyclization procedure involving a heterostilbene to establish the tricyclic CBI core, followed by introduction of an additional functionalized carbon necessary for formation of the cyclopropane (Drost, K.J.; Cava, M.P. *J. Org. Chem.* **1991**, *56*, 2240-2244) (Scheme 2).

Scheme 2



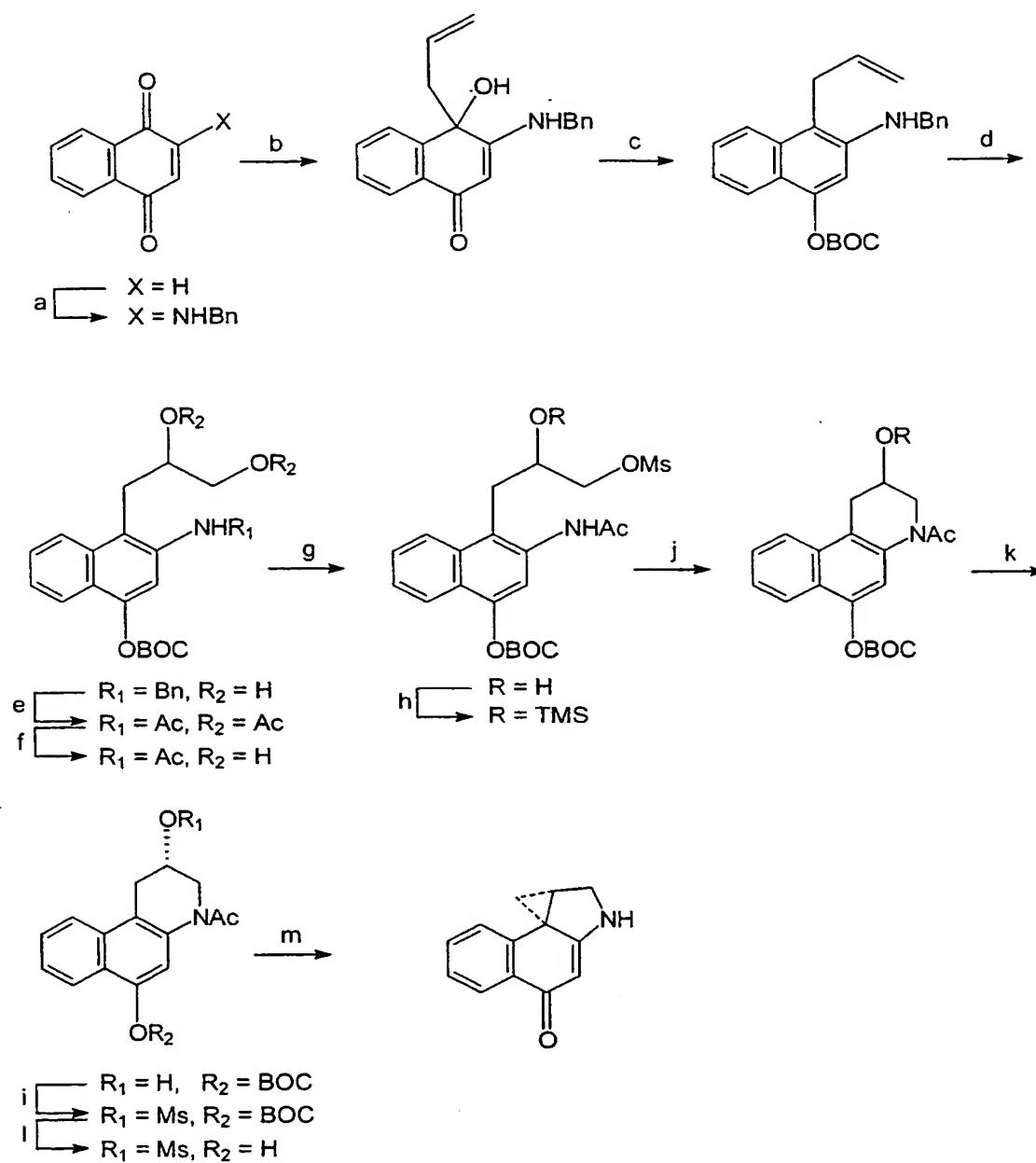
Reagents and conditions: **a**: 1) *t*-BuOK; 2) *p*-nitrobenzoic acid, Et₃N, Pd-C, hn; **b**: 1) NaBH₃CN; 2) ClCO₂CH₂CCl₃; **c**: 1) DDQ; 2) NaBH₄; **d**: Me₂NH, CH₂O; **e**: CH₃I; NaCN; **f**: NaOH; **g**: 1) MeO₂CCl, Et₃N; 2) NaBH₃CN; 3) TCBOC-Cl, Et₃N; **h**: NaOH; **i**: 1) (COCl)₂; 2) mercaptopyridine-N-oxide, DMAP, CCl₄; **j**: BCl₃·SMe₂; Et₃N

Thus, photolysis of the alkene formed from condensation of the α -methoxybenzyl phosphonate with *N*-benzylpyrrole-2-carboxaldehyde in the presence of Pd-C provided the tricyclic CBI core (Rawal, V.H.; Jones, R.J.; Cava, M.P. *J. Org. Chem.* **1987**, 52, 19-28). A four step debenzylation sequence followed by a regioselective Mannich alkylation was employed to provide the key CBI intermediate.

Aristoff synthesis of CBI (Scheme 3)

An alternative approach to the CBI subunit was described by Aristoff and coworkers (Aristoff, P.A.; Johnson, P.D.; Sun, D. *J. Med. Chem.* **1993**, 36, 1956-1963. Aristoff, P.A.; Johnson, P.D. *J. Org. Chem.* **1992**, 57, 6234-6239). 1-Allyl-2-(benzylamino)-1-hydroxydihydronaphthalenone was prepared in two steps from 1,4-naphthoquinone. Reduction and re-aromatization was accomplished by treatment with BOC₂O followed by sodium dithionite. The racemic diol was prepared by OsO₄-catalyzed dihydroxylation. Deprotection of the benzylamine, *N*- and *O*-acetylation, acetate hydrolysis, followed by selective mesylation of the primary alcohol and TMS ether protection of the secondary alcohol preceded 6-membered ring closure upon treatment with NaH. Alcohol deprotection and resolution by chromatographic separation of the diastereomeric (*R*)-*O*-acetylmandelate esters provided the optically active materials.

Scheme 3



Reagents and conditions: a: BnNH_2 ; b: Allyl MgBr; c: BOC_2O ; $\text{Na}_2\text{S}_2\text{O}_4$; d: OsO_4 , NMO; e: HCO_2H , Pd-C; Ac_2O ; f: K_2CO_3 ; g: TMSCl , pyridine; h: NaH; j: K_2CO_3 ; k: (*R*)-O-acetylmandelic acid, EDCI; Resolution;
 5 K_2CO_3 ; i: MsCl , Et_3N ; l: TFA; m: NaH

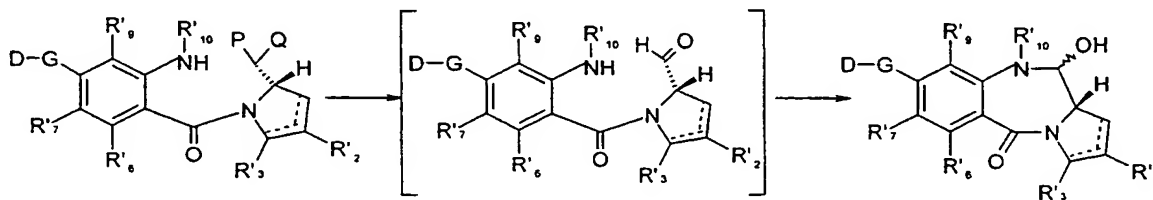
Primary alcohol activation, BOC deprotection, and transannular cyclization upon treatment with NaH provided the CBI accompanying hydrolysis of the intermediate N-Ac-CBI by water present in the
 10 reaction mixture.

Any of these three routes may be adapted to synthesise compounds according to the first aspect of the present invention, for example by starting with appropriately substituted starting
 15 materials (which may be protected), or by introducing substituents at a later stage.

Preferred Synthetic Strategies of

pyrrolo[2,1-c][1,4]benzodiazepines (also see WO 00/12506)

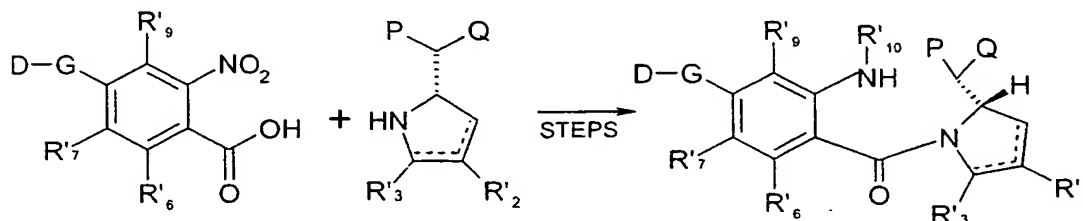
20 A key step in a preferred route to compounds corresponding to the group D(b) of the eleventh aspect of the invention or combinatorial units of formulae **XIa** and **XIb** is a cyclisation process to produce the B-ring, involving generation of an aldehyde (or functional equivalent thereof) at what will be the 11-
 25 position, and attack thereon by the pro-10-nitrogen:



In this structure, D represents XY, or a masked form thereof. The
 30 "masked aldehyde" -CPQ may be an acetal or thioacetal (possibly cyclic), in which case the cyclisation involves unmasking. Alternatively, the masked aldehyde may be an aldehyde precursor

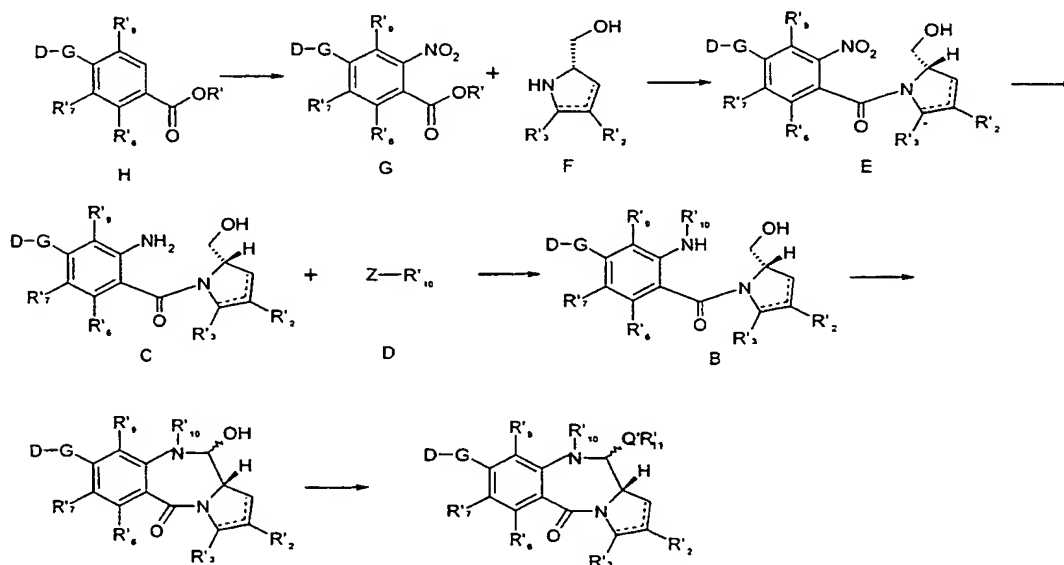
such as an alcohol -CHOH, in which case the reaction involves oxidation, e.g. by means of TPAP or DMSO (Swern oxidation).

- 5 The masked aldehyde compound can be produced by condensing a corresponding 2-substituted pyrrolidine with a 2-nitrobenzoic acid:



- 10 The nitro group can then be reduced to -NH₂ and protected by reaction with a suitable agent, e.g. a chloroformate, which provides the removable nitrogen protecting group in the desired compound.

- 15 A process involving the oxidation-cyclization procedure is illustrated in scheme 4 (an alternative type of cyclisation will be described later with reference to scheme 5).



Scheme 4

If R'₁₁ is other than hydrogen, the final compound may be prepared by direct etherification of the alcohol. Compounds with Q' = S can be prepared by treatment of the corresponding alcohol with R'₁₁SH, and a catalyst (usually a Lewis Acid such as Al₂O₃). If Q' = NH, then these compounds can be prepared by reacting the alcohol with an amine R'₁₁NH and a catalyst (usually a Lewis Acid).

Exposure of the alcohol **B** (in which the 10-nitrogen is generally protected as an amide carbamate) to tetrapropylammonium perruthenate (TPAP)/N-methylmorpholine N-oxide (NMO) over A4 sieves results in oxidation accompanied by spontaneous B-ring closure to afford the desired product. The TPAP/NMO oxidation procedure is found to be particularly convenient for small scale reactions while the use of DMSO-based oxidation methods, particularly Swern oxidation, proves superior for larger scale work (e.g. > 1 g).

The uncyclized alcohol **B** may be prepared by the addition of a nitrogen protecting reagent of formula **D**, which is preferably a chloroformate or acid chloride, to a solution of the amino alcohol **C**, generally in solution, generally in the presence of a base such as pyridine (preferably 2 equivalents) at a moderate temperature (e.g. at 0°C). Under these conditions little or no O-acylation is usually observed.

The key amino alcohol **C** may be prepared by reduction of the corresponding nitro compound **E**, by choosing a method which will leave the rest of the molecule intact. Treatment of **E** with tin (II) chloride in a suitable solvent, e.g. refluxing methanol, generally affords, after the removal of the tin salts, the desired product in high yield.

Exposure of **E** to hydrazine/Raney nickel avoids the production of tin salts and may result in a higher yield of **C**, although this method is less compatible with the range of possible C and A-ring substituents. For instance, if there is C-ring unsaturation

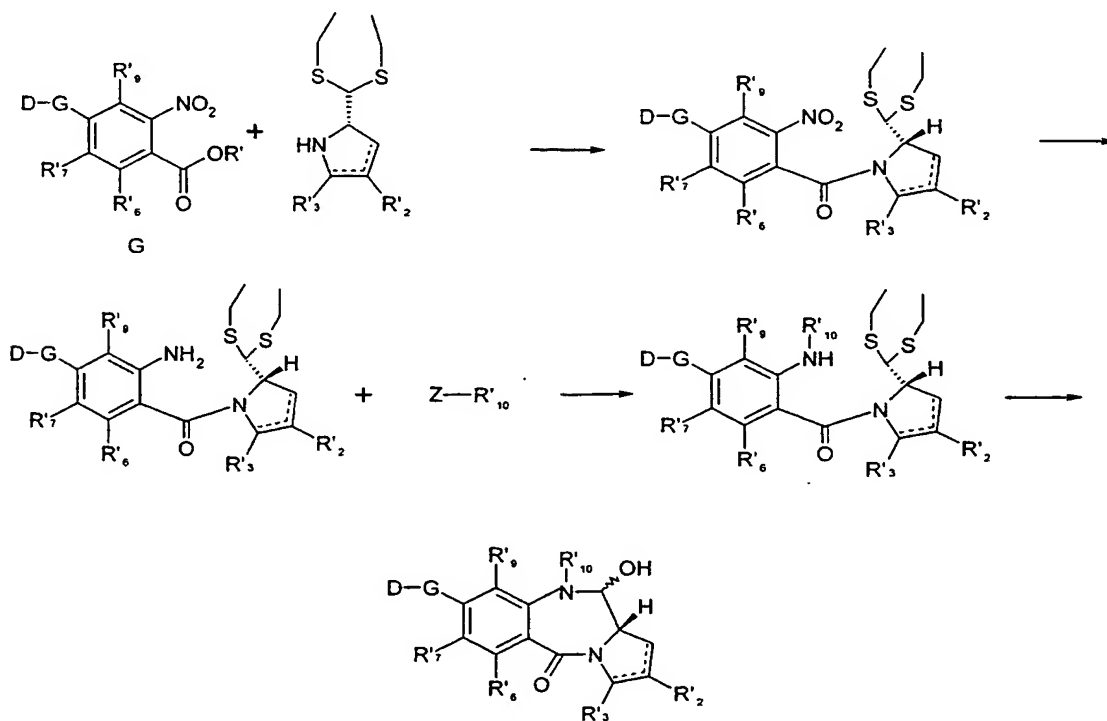
(either in the ring itself, or in R_2 or R_3), this technique may be unsuitable.

The nitro compound of formula **E** may be prepared by coupling the appropriate *o*-nitrobenzoyl chloride to a compound of formula **F**, e.g. in the presence of K_2CO_3 at $-25^\circ C$ under a N_2 atmosphere.

Compounds of formula **F** can be readily prepared, for example by olefination of the ketone derived from L-trans-4-hydroxy proline. The ketone intermediate can also be exploited by conversion to the enol triflate for use in palladium mediated coupling reactions.

The *o*-nitrobenzoyl chloride is synthesised from the *o*-nitrobenzoic acid (or alkyl ester after hydrolysis) of formula **G**, which itself is prepared from the vanillic acid (or alkyl ester) derivative **H**. Many of these are commercially available and some are disclosed in Althuis, T. H. and Hess, H. J., *J. Medicinal Chem*, 20(1), 146-266 (1977).

Alternative Cyclisation (Scheme 5)



Scheme 5

In scheme 4, the final or penultimate step was an oxidative cyclisation. An alternative approach, using thioacetal coupling/unmasking, is shown in scheme 5. Mercury-mediated unmasking causes cyclisation to the desired compound.

The thioacetal intermediates may be prepared as shown in scheme 2: the thioacetal protected C-ring [prepared via a literature method: Langley, D.R. & Thurston, D.E., *J. Organic Chemistry*, **52**, 91-97 (1987)] is coupled to the *o*-nitrobenzoic acid (or alkyl ester after hydrolysis) **G** using a literature procedure. The resulting nitro compound cannot be reduced by hydrogenation, because of the thioacetal group, so the tin(II) chloride method is used to afford the amine. This is then N-protected, e.g., by reaction with a chloroformate or acid chloride, such as *p*-nitrobenzylchloroformate.

Acetal-containing C-rings can be used as an alternative in this type of route with deprotection involving other methods including the use of acidic, or perhaps Lewis Acid, conditions.

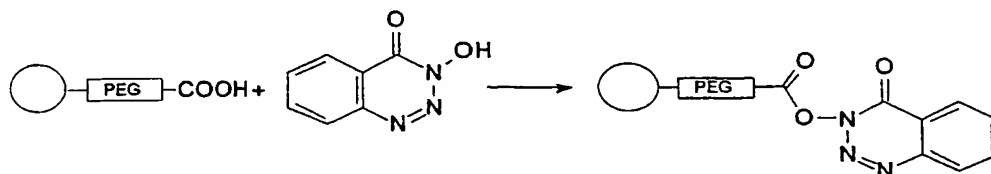
In the above synthesis schemes, the derivatisation of the A-ring is shown as being complete before the compounds are attached to the solid support. This is preferred if the substituents are groups such as alkoxy or nitro. On the other hand, substituent groups such as alkyl or alkenyl could be added to the A-ring after the coupling of the compound to the solid support. This may be achieved by R'₆, R'₇, or R'₈ being easily replaceable groups, such as a halogen atom.

An alternative synthesis approach to those detailed above is to protect the Pro N10 position on the component which will form the A-ring before joining the component which will form the C-ring.

Embodiments of the present invention will now be described by way of example.

Example 1: CBI bound to TG-Carboxy resin

Activation of TG-Carboxy resin with 3-hydroxy-1,2,3-benzotriazin-4(3H)-one (HO-Dhbt)

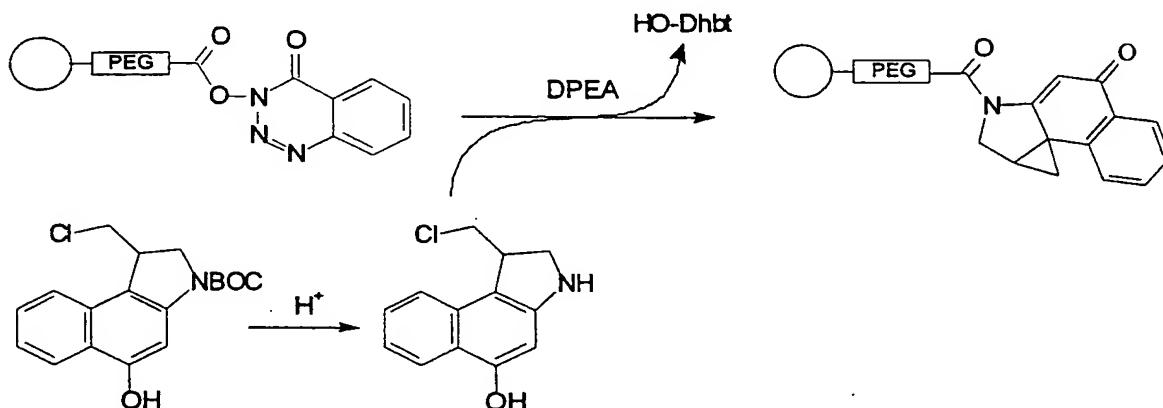


10 TG Carboxy resin (Nova-Biochem, 80 mg 0.020 mmol) was suspended and swollen in DMF (1 mL) for 10 minutes. DMF was removed by suction and the resin treated with diisopropyl carbodiimide (DIC) (34 mmL, 10 eq) in DMF (1 mL). The mixture was shaken for 15 minutes at -15°C (acetone-ice bath), then rinsed with DMF (2 x 1 mL).

15 The resin was suspended in DMF (1 mL), cooled to -10°C and treated with HO-Dhbt (33 mg, 10 eq). The mixture was shaken for 30 minutes at -10°C, followed by 4 hours shaking at 0°C and was then allowed to stand at 0°C overnight.

20 The reaction mixture was filtered and the resin washed with DMF (2 x 1 mL), dichloromethane (2 x 1 mL), MeOH (2 x 1 mL) and diethyl ether (1 mL). The activated resin was stored in freezer until ready for use.

Coupling of O-Dhbt activated TG-Carboxy resin with seco-CBI, in presence of DIPEA

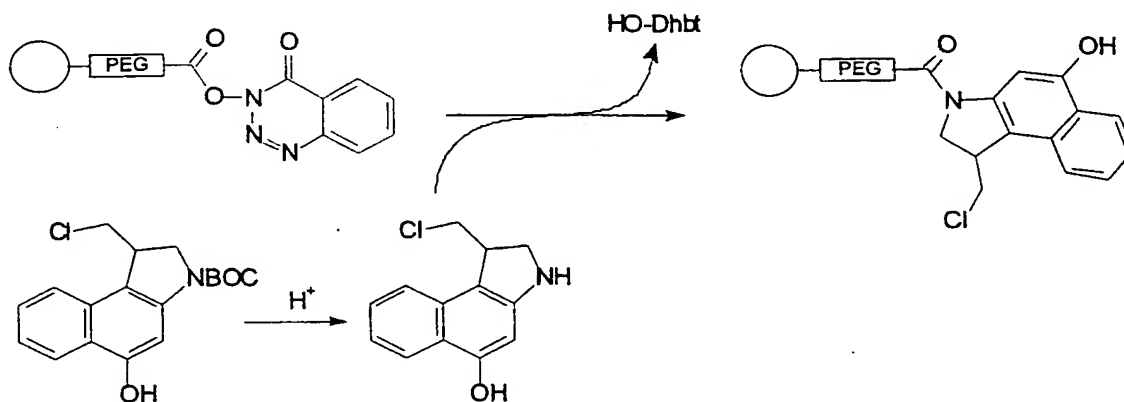


5 Seco-CBI has a very limited storage time, and for this reason BOC-seco-CBI is deprotected immediately before coupling to the activated resin. BOC-seco-CBI (20 mg 0.06 mmol) was dissolved in HCl in anhydrous ethyl acetate (3M, 2.8 mL). The mixture was kept at 0°C for 30 minutes and then allowed to warm to room temperature over 30 minutes. TLC (4:1 petroleum ether:diethyl ether) revealed complete reaction; the reaction mixture was evaporated under reduced pressure and stored at -20°C.

15 O-Dhbt activated resin (40 mg) was treated with seco-CBI dissolved in DMF (1 mL). DIPEA (3 eq, 7.7 mg, 10 mmL) was added to the suspension and the mixture was shaken for 30 minutes at room temperature. The reaction mixture was filtered and the resin was washed with DMF (1 mL), dichloromethane (1 mL), methanol (1 mL) and diethyl ether (2 x 1 mL). After thorough evaporation of solvents the resin was stored at -20°C.

Example 2: seco-CBI bound to TG-Carboxy resin

Coupling of O-Dhbt activated TG-Carboxy resin with seco-CBI, in absence of base



BOC-seco-CBI (20 mg, 0.06 mmol) was dissolved in HCl solution in anhydrous ethyl acetate (3M, 2.8 mL). The mixture was kept at 0°C for 30 minutes and then allowed to warm to room temperature over 30 minutes. TLC (4:1 = petroleum ether:diethyl ether) revealed complete reaction; the reaction mixture was evaporated at reduced pressure and directly used for the coupling without further purification.

O-Dhbt-activated resin (see Example 1) (40 mg, 0.01 mmol) was treated with seco-CBI dissolved in DMF (1 mL); the mixture was agitated for 1 hour at room temperature. The reaction mixture was then filtered and the resin was washed with DMF (1 mL), dichloromethane (1 mL), methanol (1 mL) and diethyl ether (2x1 mL). After thorough evaporation of excess solvent the resin was stored at -20°C.

Example 3 - Synthesis of a CBI - hexapeptide libraryGeneral Procedures*General procedure for acetylation/endcapping*

After each coupling step the resin was treated with a mixture of pyridine (30%) and acetic anhydride (20%) in dichloromethane, to acetylate any free amino groups that had not been coupled to an amino acid. In this way the formation of undesirable oligopeptides (carrying less than the expected number of amino acids) could be avoided.

The resin was treated with the acetylating reagent (3 mL) and the slurry was agitated at room temperature for 1 hour. The reagents were then removed by filtration and the resin was rinsed with dichloromethane (2 x 5 mL) and methanol (2 x 5 mL).

General procedure for Fmoc deprotection

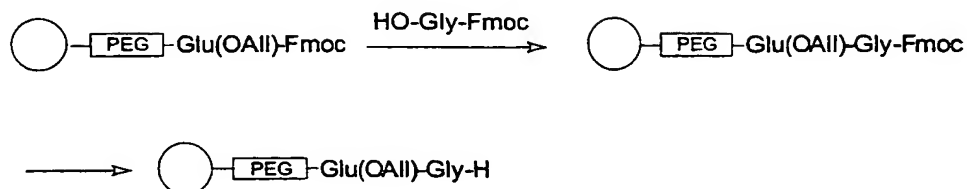
The resin was treated with a solution of piperidine in DMF (20%, 3 mL). The mixture was then agitated for 2 hours at room temperature. Excess solvent was then removed by suction and the resin was rinsed with DMF (2 x 5 mL), dichloromethane (2 x 5 mL) and methanol (2 x 5 mL).

Library Generation*Fmoc-Glu-OAll coupled to NovaSyn TG resin*

NovaSyn TG amino resin (0.345 g, load 0.29 mmol/g, 0.1 mmol) was suspended and swollen in DMF (2 mL), under agitation (1000 rpm) for 30 minutes. Excess solvent was then removed and a solution of Fmoc-Glu-OAll (0.123g, 0.3 mmol, 3 eq), TBTU (0.096 g, 0.3 mmol, 3 eq) and DIPEA (0.052 mL, 0.3 mmol, 3 eq) in DMF (3 mL) was added to the swollen resin. The resulting mixture was agitated at 1000 rpm at room temperature, overnight. The reaction mixture was filtered and the resin rinsed with DMF (2 x 2 mL), dichloromethane (2 x 5 mL) and methanol (2 x 5 mL).

The resin was then acetylated (see General procedure for acetylation) and deprotected (see General procedure for Fmoc deprotection).

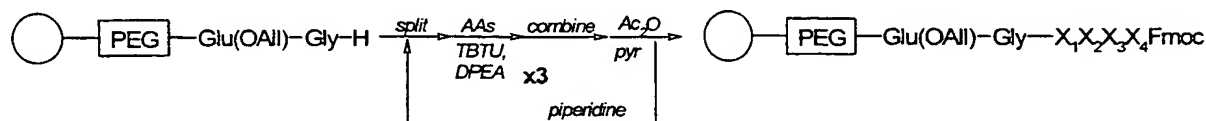
HO-Gly-Fmoc coupled to P-Glu(OAll)-H



5 The resin was suspended in DMF (2 mL) and allowed to swell for 30 minutes at room temperature, accompanied by agitation (1000 rpm). Excess DMF was removed and a solution of Fmoc-Gly-OH (0.089 g, 0.3 mmol, 3 eq), TBTU (0.096 g, 0.3 mmol, 3 eq) and DIPEA (0.052 mL, 0.3 mmol, 3 eq) in DMF (3 mL) was added to the resin. The mixture
 10 was allowed to shake at room temperature for 12 hours. Excess reagents were then removed by filtration and the resin was rinsed with DMF (2 x 5 mL), dichloromethane (2 x 5 mL) and methanol (2 x 5 mL). The resin was then acetylated (see General procedure for acetylation) and deprotected (see General procedure for Fmoc
 15 deprotection) to afford the resin-bound dipeptide P-Glu(OAll)-Gly-H.

Split & Mix procedure for the resin bound hexapeptide

P-Glu(OAll)-Gly- $X_1X_2X_3X_4$ -H sublibrary



25 The resin was suspended in 3:1 mixture of 1,2-dichloroethane (DCE) and DMF and equally partitioned into 17 4 mL Alltech tubes. Each tube thus contained $0.1/17 \text{ mmol} = 5.88 \times 10^{-6} \text{ mol}$ of resin-bound dipeptide. Excess solvent was removed *in vacuo*, and the resin was suspended in DMF (200 mL) and agitated for 30 minutes. The 17 amino acids ($1.76 \times 10^{-5} \text{ mmol}$, 3 eq for each step, $7.04 \times 10^{-5} \text{ mmol}$ for 4 steps) were weighed into 17 vials:

- | | | |
|----|-------------------|-------|
| 1. | Fmoc-Ala-OH | 22 mg |
| 2. | Fmoc-Asn-OH | 25 mg |
| 3. | Fmoc-Asp(OtBu)-OH | 29 mg |

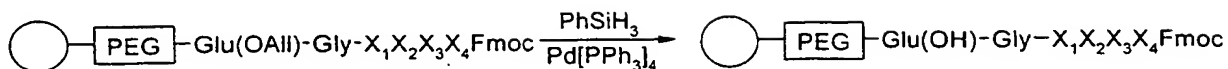
40

	4.	Fmoc-Gln-OH	26 mg
	5.	Fmoc-Glu(OtBu)-OH	30 mg
	6.	Fmoc-Gly-OH	21 mg
	7.	Fmoc-Ile-OH	25 mg
5	8.	Fmoc-Leu-OH	25 mg
	9.	Fmoc-Lys(BOC)-OH	33 mg
	10.	Fmoc-Met-OH	26 mg
	11.	Fmoc-Phe-OH	27 mg
	12.	Fmoc-Pro-OH	24 mg
10	13.	Fmoc-Ser(tBu)-OH	27 mg
	14.	Fmoc-Thr(tBu)-OH	28 mg
	15.	Fmoc-Trp(BOC)-OH	37 mg
	16.	Fmoc-Tyr(tBu)-OH	32 mg
	17.	Fmoc-Val-OH	24 mg

Each amino acid was dissolved in DMF (2 mL); an aliquot of each solution (0.5 mL, corresponding to $1.76 \cdot 10^{-5}$ mmol, 3 eq of each amino acid) was added to the appropriate tube. TBTU ($1.76 \cdot 10^{-5}$ mmol \times 17 = $2.99 \cdot 10^{-4}$, 96 mg) and DIPEA ($1.76 \cdot 10^{-5}$ mmol \times 17 = $2.99 \cdot 10^{-4}$, 52 mL) were separately dissolved in DMF (1.7 mL) and each solution was evenly distributed, delivering 3 eq of each reagent, to each one of the 17 tubes.

The reaction tubes were agitated at room temperature for 12 hours, then the reagents and solvents were removed *in vacuo* and the resin was rinsed with DMF (2 x 1 mL each tube), DCM (2 x 1 mL each tube) and methanol (2 x 1 mL each tube). The resin was then suspended in 3:1 mixture of 1,2-dichloroethane and DMF and recombined. The recombined resin was acetylated (3 mL of acetylating reagent, 1 hour, room temperature) and deprotected (3 mL of 20% piperidine in DMF, 2 hours, room temperature).

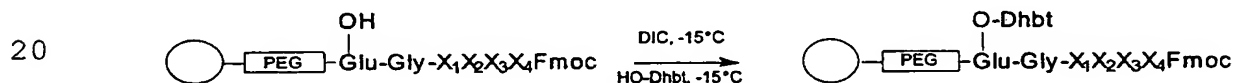
The procedure was repeated 3 more times. At the end of the 4th amino acid coupling the deprotection step was not executed.

Deprotection of Allyl ester

5 The resin was suspended in DCM (2 mL), phenylsilane (2.4 mmol, 24 eq, 0.29 mL) in DCM (1 mL) of was added and the mixture was shaken at room temperature for 10 minutes.

10 A catalytic amount of $\text{Pd}(\text{PPh}_3)_4$ (0.01 mmol, 0.1 eq, 11 mg) in DCM (0.5 mL) was added and the reaction mixture was shaken for further 10 minutes.

15 The reagents were filtered and the resin was rinsed with DCM (2 x 5 mL) and methanol (2 x 5 mL). The procedure was repeated once again and the resin was finally dried under reduced pressure.

Activation of resin with 3-hydroxy-1,2,3-benzotriazin-4(3H)-one (HO-Dhbt)

20 The resin was suspended and swollen in DMF (2 mL) for 10 minutes. Excess DMF was removed and the resin was treated with a solution of diisopropyl carbodiimide (DIC) (156 mL, 1 mmol, 10 eq) in DMF (2 mL). The resulting slurry was shaken for 15 minutes at -15°C (acetone-ice bath), then washed with DMF (2 x 2 mL).

30 The resin was resuspended in DMF (2 mL), cooled to -10°C and treated with HO-Dhbt (163 mg 1 mmol, 10 eq). The mixture was shaken for 30 minutes at -10°C , followed by 4 hours at 0°C and then allowed to stand at 0°C overnight.

35 The reaction mixture was filtered and the resin washed with DMF (2 x 2 mL), dichloromethane (2 x 2 mL), MeOH (2 x 2 mL) and diethyl ether (2 mL). The activated resin was stored in freezer until required for use.

Coupling of O-Dhbt activated resin with seco-CBI, in presence of DIPEA



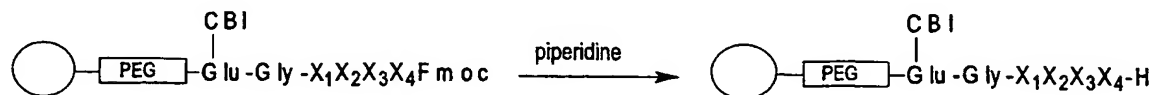
BOC-seco-CBI (see example 2) was deprotected immediately before coupling to the activated resin. BOC-seco-CBI (100 mg, 0.3 mmol) was dissolved in a solution of HCl in anhydrous ethyl acetate (3M, 15 mL). The mixture was cooled at 0°C for 30 minutes and then allowed to warm to room temperature over 30 minutes. TLC (4:1 = petroleum ether:diethyl ether) revealed complete reaction; the reaction mixture was evaporated under reduced pressure and stored at -20°C.

O-Dhbt-activated resin was treated with seco-CBI dissolved in DMF (2 mL) and the mixture was shaken for 60 minutes at room temperature. The reaction mixture was filtered and the resin was washed with DMF (2 x 2 mL), dichloromethane (2 x 2 mL), methanol (2 x 2 mL) and diethyl ether (2 x 1 mL). The resin was dried under reduced pressure and stored in a cool, dry place.

t-Butyl- and BOC- deprotection on side chains

The resin was treated with a solution of TIS/TFA in dichloromethane (2%, 2 mL). The mixture was shaken for 2 hours at room temperature. Excess reagents were removed *in vacuo* and the resin was rinsed with dichloromethane (2 x 5 mL) and methanol (2 x 5 mL).

Fmoc deprotection



The resin was treated with a solution of piperidine in DMF (20%, 3 mL), and agitated at room temperature for 2 hours. The mixture was filtered and the resin was washed with DMF (2 x 5 mL),

dichloromethane (2 x 5 mL) and methanol (2 x 5 mL). The resin was then dried under reduced pressure.

Screening

- 5 The resulting library was screened against rhodamine labelled double stranded DNA with the sequence Label-5'-GCG TAA AAA CGC = 3'.

10 The sublibrary was mixed with the DNA sequence (5 pmol/mL) and incubated at 37°C for 24 hours with occasional mixing. After 24 hours, the sublibrary was washed 4 times with TE buffer pH 7.6 or PBS. To identify the beads to which most labelled DNA had bound, agarose gel slides were prepared as follows. ~ 500 mL of 0.25% sea plaque agarose was layered onto a clean transparent slide and
15 allowed to cool and set. The incubated beads were then mixed with another - 500 mL of 0.25% sea plaque agarose solution and layered onto the precoated slides, and allowed to cool and set.

20 The reddest beads were identified by eye under a dissecting light microscope, and then retrieved by adding ~ 1 mL of water to dried agarose slide to enable their removal using a p10 gilson pipette with a fine tip. The removed beads were then placed into a 1 mL Eppendorf PCR tube ready for identification.

Identification

25 The identification of the sequences of the most active compounds was carried out using automated Edman degradation and reversed-phase HPLC.

30 Pulsed liquid-phase N-terminal sequencing was performed using an Applied Biosystems (ABI) 477A automatic protein sequencer. The selected labelled beads were loaded onto a glass fibre disc which had previously been pre-cycled once. The disc was placed in the sequencer and pre-cycled once, then six cycles of Edman
35 degradation were performed (Edman, P and Begg, G (1967) Eur. J. Biochem. 1, 80). The released phenylthiohydantion (PTH-) amino acid derivatives were identified by reversed-phase HPLC analysis.

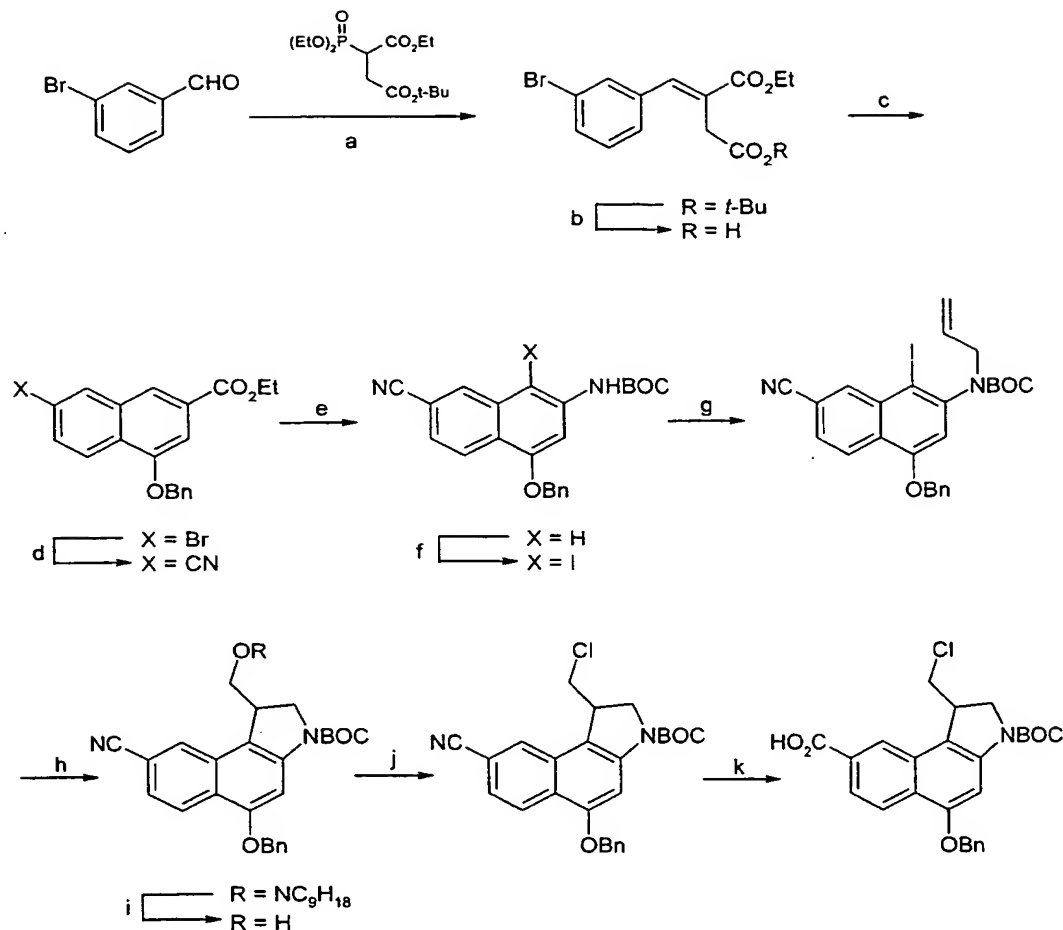
The four most active compounds were those with the following sequences:

	CBI-QGVKKK	CBI-QGLVAG
5	CBI-QGNKKA	CBI-QGQKNS

Example 4: Synthesis of A CBI Combinatorial Building Block: 7-carboxy-1,2,9,9a-tetrahydrocyclopropa[c]benzo[e]indol-4-one

- 10 The synthesis proceeds from a modified Stobbe condensation/Friedel-Crafts acylation for generation of the functionalised precursor, followed by 5-exo-trig aryl radical-alkene cyclization.
- 15 Wadsworth-Horner-Emmons condensation of 3-bromo-benzaldehyde with the Sargent phosphonate predominantly provides the E-isomer, which in turn undergoes acid-catalyzed deprotection and Friedel-Crafts acylation (Scheme 6).

Scheme 6



Reagents and conditions: **a:** NaH, Sargent phosphonate; **b:** TFA; **c:** 1) $\text{Ac}_2\text{O-KOAc}$; 2) K_2CO_3 ; 3) BnBr , K_2CO_3 ; **d:** CuCN ; **e:** 1) LiOH ; 2) DPPA , $t\text{-BuOH}$; **f:** NIS ; **g:** allyl Br, NaH; **h:** Bu_3SnH , TEMPO; **i:** Zn-HOAc ; **j:** $\text{Ph}_3\text{P-CCl}_4$; **k:** NaOH

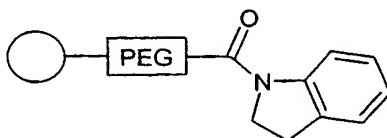
Aromatic nucleophilic substitution, ester hydrolysis and Curtius rearrangement effected by treatment with DPPA are followed by regioselective C4 iodination and N-alkylation with allyl bromide. The aryl radical-alkene cyclization by means of TEMPO as radical trap, as described in Boger synthesis of CBI, provides the tricyclic system that, after conversion to the primary chloride and base-catalyzed hydrolysis of the cyano group, should give the desired building block for a combinatorial library.

Example 5: Screening of CBI on bead against DNA sequence

The CBI on bead synthesised in example 1 was screened against rhodamine labelled double stranded DNA with the sequence:

5 Label-5'-GCG TAA AAA CGC-3'.

For comparison, indoline was added to O-Dhbt-activated TG Carboxy resin (as prepared in example 1) to yield a comparative resin that does not covalently interact with DNA.



The resin alone was also tested to provide a background measurement. The screening protocol used was:

1. Weigh out approximately 1 mg of resin into an Eppendorf.
2. Incubate the resin with annealed rhodamine labelled double strand DNA for 24 hours at 37°C.
3. After 24 h incubation wash the resin 4 times and re-suspend beads in 50 mL of TE (EDTA and Tris buffer) or PBS (phosphate buffered saline).
4. Resin was transferred to a black 96 well plate with transparent base, fluorescence was measured from below at 635 nm using a Tecan Spectrofluor (590 nm excitation, 635 nm emission).

Resin	Relative Fluorescence Units
TG Carboxy Resin	892
Indoline Bound to Resin	625
CBI Bound to Resin	2947

Only a small amount of DNA binding was observed with the acid resin and with the resin loaded with an inactive indoline CBI mimic. An almost five fold differential was observed between the CBI resin and Indoline resin demonstrating DNA binding by the resin bound CBI. It is anticipated that a higher concentration of DNA would lead to a higher differential between the resins as it

is possible that the CBI-resin is not saturated with DNA at this concentration.

Example 6: Synthesis of a CBI - PBD - hexapeptide library

General procedure for acetylation/endcapping

After any coupling step the resin was treated with a mixture of pyridine (30%) and acetic anhydride (20%) in dichloromethane, to acetylate any free amine that has not been coupled to an amino acid. The formation of unexpected oligopeptides (carrying less than the expected number of aminoacids) is avoided.

The resin was treated with 3 mL of acetylating reagent and the slurry was agitated at room temperature for 1 hour. The reagents were then removed by filtration and the resin was rinsed with dichloromethane (2 x 5 mL) and methanol (2 x 5 mL).

General procedure for Fmoc deprotection

The resin was treated with 3 mL of a solution of piperidine (20%) in DMF. The mixture was then agitated for 2 hours at room temperature. The solvent was filtered off and the resin was rinsed with DMF (2 x 5 mL), dichloromethane (2 x 5 mL) and methanol (2 x 5 mL).

Fmoc-Glu-OAll coupled to NovaSyn TG resin



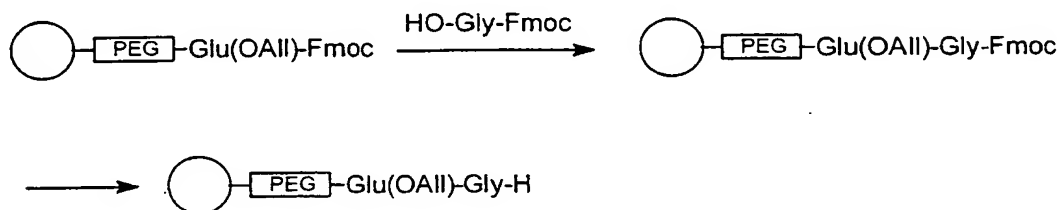
NovaSyn TG amino resin 130 μ (0.345 g, load 0.29 mmol/g, 0.1 mmol) was suspended and swollen into 2 mL of DMF, under agitation (1000 rpm) for 30 minutes. The solvent was then removed and a solution of Fmoc-Glu-OAll (0.123g, 0.3 mmol, 3 eq), TBTU (0.096g, 0.3 mmol, 3 eq) and DIPEA (0.052 mL, 0.3 mmol, 3 eq) in 3 mL of DMF was added to the swollen resin. The resulting mixture was agitated at 1000 rpm, room temperature, overnight. The reaction mixture was

filtered and the resin rinsed with DMF (2 x 2 mL), dichloromethane (2 x 5 mL) and methanol (2 x 5 mL).

- 5 The resin was then acetylated (see General procedure for acetylation) and deprotected (see General procedure for Fmoc deprotection).

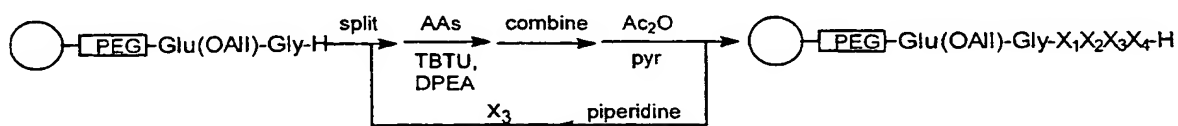
HO-Gly-Fmoc coupled to P-Glu(OAll)-H

10



- 15 The resin was suspended in 2 mL of DMF and was allowed to swell for 30 minutes at room temperature, whilst being agitated (1000 rpm). DMF was removed and a solution of Fmoc-Gly-OH (0.089 g, 0.3 mmol, 3 eq), TBTU (0.096 g, 0.3 mmol, 3 eq) and DIPEA (0.052 mL, 0.3 mmol, 3 eq) in 3 mL of DMF was added to the resin. The mixture was shaken at room temperature for 12 hours. The reagents were then removed by filtration and the resin was rinsed with DMF (2 x 5 mL), dichloromethane (2 x 5 mL) and methanol (2 x 5 mL). The resin was then acetylated (see General procedure for acetylation) and deprotected (see General procedure for Fmoc deprotection) to afford the resin-bound dipeptide P-Glu(OAll)-Gly-H.

- 25 Split & Mix procedure for the resin bound hexapeptide P-Glu(OAll)-Gly-X₁X₂X₃X₄-H sublibrary



30

The resin was suspended in 3:1 mixture of 1,2-dichloroethane and DMF and equally partitioned into 17 4 mL Alltech tubes. Each tube thus contained 0.1/17 mmol = 5.88 10⁻⁶ mol of resin-bound dipeptide. Solvent was removed *in vacuo*, and the resin was

suspended in 200 mL of DMF and agitated for 30 minutes. The 17 aminoacids ($1.76 \cdot 10^{-5}$ mmol, 3 eq for each step, $7.04 \cdot 10^{-5}$ mmol for 4 steps) were weighed into 17 vials:

5	1.	Fmoc-Ala-OH	22 mg
	2.	Fmoc-Asn-OH	25 mg
	3.	Fmoc-Asp(OtBu)-OH	29 mg
	4.	Fmoc-Gln-OH	26 mg
	5.	Fmoc-Glu(OtBu)-OH	30 mg
10	6.	Fmoc-Gly-OH	21 mg
	7.	Fmoc-Ile-OH	25 mg
	8.	Fmoc-Leu-OH	25 mg
	9.	Fmoc-Lys(BOC)-OH	33 mg
	10.	Fmoc-Met-OH	26 mg
15	11.	Fmoc-Phe-OH	27 mg
	12.	Fmoc-Pro-OH	24 mg
	13.	Fmoc-Ser(tBu)-OH	27 mg
	14.	Fmoc-Thr(tBu)-OH	28 mg
	15.	Fmoc-Trp(BOC)-OH	37 mg
20	16.	Fmoc-Tyr(tBu)-OH	32 mg
	17.	Fmoc-Val-OH	24 mg

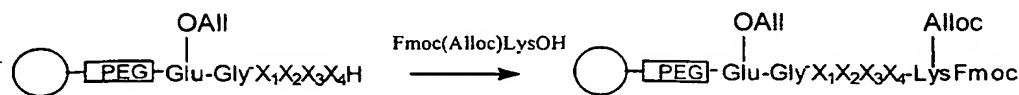
Each amino acid was dissolved in 2 mL of DMF; 0.5 mL of this solution (corresponding to $1.76 \cdot 10^{-5}$ mmol, 3 eq of each amino acid) was added to the appropriate tube. TBTU ($1.76 \cdot 10^{-5}$ mmol \times 17 = $2.99 \cdot 10^{-4}$, 96 mg) and DIPEA ($1.76 \cdot 10^{-5}$ mmol \times 17 = $2.99 \cdot 10^{-4}$, 52 mL) were separately dissolved in 1.7 mL of DMF and each solution was evenly partitioned, addressing 3 eq of each reagent, to each one of the 17 tubes.

The reaction tubes were agitated at room temperature for 12 hours, then the reagents and solvents were removed *in vacuo* and the resin was rinsed with DMF (2 x 1 mL each tube), DCM (2 x 1 mL each tube) and methanol (2 x 1 mL each tube). The resin was then suspended in 3:1 mixture of 1,2-dichloroethane and DMF and recombined. The recombined resin was acetylated (3 mL of acetylating reagent, 1 hour, room temperature) and deprotected (3 mL of 20% piperidine in DMF, 2 hours, room temperature).

The procedure was repeated 3 more times to afford the deprotected resin-bound hexapeptide G(OAll)EX₁X₂X₃X₄H.

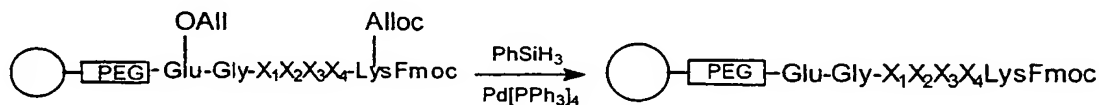
Coupling of Fmoc(Alloc)Lys-OH to resin-bound hexapeptide

5



The resin was suspended in 2 mL of DMF and was allowed to swell for 10 minutes at room temperature, whilst being agitated (1000 rpm). DMF was removed and a solution of Fmoc(Alloc)Lys-OH (0.135g, 0.3 mmol, 3 eq), TBTU (0.096g, 0.3 mmol, 3 eq) and DIPEA (0.052 mL, 0.3 mmol, 3 eq) in 3 mL of DMF was added to the resin. The mixture was shaken at room temperature for 12 hours. The reagents were then removed by filtration and the resin was rinsed with DMF (2 x 5 mL), dichloromethane (2 x 5 mL) and methanol (2 x 5 mL). The resin was then acetylated (see General procedure for acetylation) to afford the resin-bound heptapeptide.

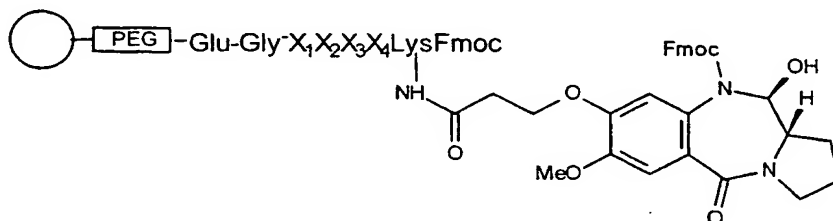
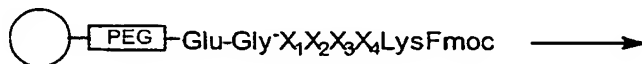
Deprotection of Alloc ester



The resin was suspended in 2 mL of DCM. Phenylsilane (2.4 mmol, 24 eq, 0.29 mL) in 1 mL of DCM was added and the mixture was shaken at room temperature for 10 minutes. Pd[PPh₃]₄ (0.01 mmol, 0.1 eq, 11 mg) in 0.5 mL of DCM was added and the reaction mixture was shaken for further 10 minutes.

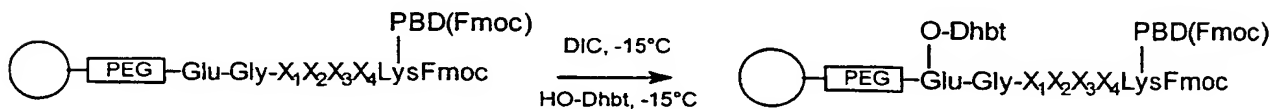
The reagents were filtered and the resin was rinsed with DCM (2 x 5 mL) and methanol (2 x 5 mL), then the procedure was repeated once again. The resin was finally dried under reduced pressure.

Coupling of N10-Fmoc protected PBD to lysine residue of the resin-bound heptapeptide



The resin was suspended in 2 mL of DMF and agitated for 10 minutes to allow swelling. DMF was then removed. A solution of N10-Fmoc-PBD acid (0.556 g, 1 mmol, 10 eq.), TBTU (0.353 g, 1.1 mmol, 1.1 eq.) and DIPEA (0.19 mL, 1.1 mmol, 1.1 eq.) in 3 mL of DMF was stirred for 30 minutes, and added to the resin. The mixture was agitated at room temperature for 12 hours, then the reagents were removed under reduced pressure and the resin was rinsed with DMF (2 x 5 mL), dichloromethane (2 x 5 mL), methanol (2 x 5 mL) and diethyl ether (1 x 2 mL).

Activation of resin with 3-hydroxy-1,2,3-benzotriazin-4(3H)-one (HO-Dhbt)



The resin was suspended and swollen in 2 mL of DMF for 10 minutes. DMF was removed and the resin was treated with 156 mL (1 mmol, 10 eq) of diisopropyl carbodiimide (DIC) in 2 mL of DMF. The resulting slurry was shaken for 15 minutes at -15°C (acetone-ice bath), then washed with DMF (2 x 2 mL).

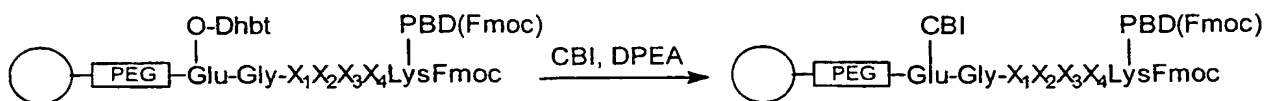
The resin was re-suspended in 2 mL of DMF, cooled at -10°C and treated with 163 mg (1 mmol, 10 eq) of HO-Dhbt. The mixture was shaken for 30 minutes at -10°C , for 4 hours at 0°C and then allowed to stand at 0°C overnight.

5

The reaction mixture was filtered and the resin washed with DMF (2 x 2 mL), dichloromethane (2 x 2 mL), MeOH (2 x 2 mL) and diethyl ether (2 mL). The activated resin was stored in freezer until ready for use.

10

Coupling of O-Dhbt activated resin with seco-CBI, in presence of DIPEA



BOC-seco-CBI was deprotected immediately before coupling to the activated resin. 100 mg (0.3 mmol) of BOC-seco-CBI was dissolved in 15 mL of 3M HCl solution in anhydrous ethyl acetate. The mixture was cooled at 0°C for 30 minutes and then allowed to warm to room temperature for further 30 minutes. TLC (4:1 = petroleum ether:diethyl ether) revealed reaction completion; the reaction mixture was evaporated at reduced pressure and stored at -20°C .

25

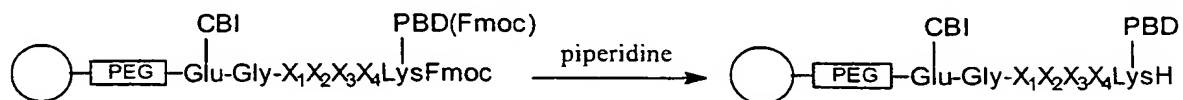
O-Dhbt-activated resin was treated with seco-CBI dissolved in 2 mL of DMF and the mixture was shaken for 60 minutes at room temperature. At the end the reaction mixture was filtered and the resin was washed with DMF (2 x 2 mL), dichloromethane (2 x 2 mL), methanol (2 x 2 mL) and diethyl ether (2 x 1 mL). The resin was dried under reduced pressure and stored in a cool, dry place.

30

20220220 20269001

t-Butyl- and BOC- deprotection on side chains

The resin was treated with a solution of TIS/TFA 2% in dichloromethane (2 mL). The mixture was shaken for 2 hours at room temperature. The reagents were removed *in vacuo* and the resin was rinsed with dichloromethane (2 x 5 mL) and methanol (2 x 5 mL).

Fmoc deprotection

The resin was treated with 3 mL of a solution of piperidine (20%) in DMF, and agitated at room temperature for 2 hours. At the end the mixture was filtered and the resin was washed with DMF (2 x 5 mL), dichloromethane (2 x 5 mL) and methanol (2 x 5 mL). The resin was then dried under reduced pressure.

202220"2026900T

15